Design for Calibrated Trust for Acceptance of Autonomous Vehicles

Master Thesis
Strategic Product Design
David Callisto Valentine

Author

David Callisto Valentine
d.c.valentine@student.tudelft.nl

Master Thesis

MSc. Strategic Product Design
Faculty of Industrial Design Engineering
Delft University of Technology

Graduation Committee

Chair : **Dr. Kim, E.Y.**
Assistant Professor I Design for Dynamic Stability
Department of Design, Organization, and Strategy (DOS)

Mentor : **Ir. Smit, I.R.**
Visiting Professor and PhD candidate I Lab Director Cities of Things Lab
Human Information Communication Design, Human-Centered Design

Design Lab

Cities of Things Delft Design Lab
People in Transit

Design for Calibrated Trust for acceptance of autonomous vehicles

Master Thesis
by
David Callisto Valentine

Preface

Dear Reader,

The report is the culmination of two years of my journey from a Mechanical Engineer to a Strategic Designer, in the form of a 20 week graduation project. During these 2 years I have had the pleasure to meet and interact with a diverse diaspora of designers, entrepreneurs, innovators, engineers and most importantly lovely human beings. I would thus like to take a few column inches within this report to acknowledge their contribution to this journey.

I would like to start out by thanking my supervisory team, Dr. Euiyoung Kim and Ir. Iskander Smit. Thank you for all the feedback during our meetings, the insightful opinions and resources, the support during the transition of the project to meet the modified working conditions during the COVID-19. Also, the constant push and encouragement to better the project and my practices as a designer.

Secondly, I would like to thank my family, friends and my teammates during the various projects for the support and the numerous learnings. A special thanks to Devesh and Anirban for the weekly calls. Ishit, Dheebak and Sid for the numerous Fortnite games we spent playing during the lockdown. My Design Strategy Project Team, Yael, Fiona and Alejandra for being supportive during the initial phases of formulating my graduation brief. My brother and parents for the constant support during these two years and especially during the pandemic.

A special thanks to all the experts and other collaborators that took time out to be a part of the project, either in the data collection phase/testing and validation phase. The insights and recommendations were insightful in shaping the project and the final outcome.

I would also like to thank TU Delft, Dutch Government and all health care workers for making it possible to continue to work and study during the pandemic. I apologise if I have missed thanking someone, and hope you enjoy reading the report.

Regards,
David Callisto Valentine

Executive Summary

The advent of a society in which autonomous technology coexists with humans is an inevitability. The project focuses on one such autonomous technology in the form of autonomous vehicles or self driving cars. The benefits of such automation is well documented in academia and is supported by the investment by some of the biggest automobile and technology manufacturers in the research and development of autonomous vehicles. However, there exist certain challenges in realizing the full potential of autonomous vehicles. One such challenge is the attribute of trusting an autonomous vehicle. The project looks at the idea of trust in automation and dives deeper into the concept of calibrated trust as an approach to designing autonomous vehicles for increased acceptance of autonomous vehicles. The project is conducted in association with the Cities of Things Design Lab and People in Transit.

Calibrated trust is defined as the balance between the capabilities of autonomous vehicles to the expectations of the end user. In essence it is the creation of an appropriate mental model by the end user. Through literature research and qualitative analysis, prominent challenges in achieving calibration were identified as: approach to designing for socio-technical systems, misalignment in communication between stakeholders, product branding and customer experience. Since, the focus of the project was towards the design and development team, the first two challenges i.e. approach to designing for socio-technical systems and misalignment in communication between stakeholders were selected to define the final design question and direction.

The final design intervention is a Calibrated Trust Toolkit that can be used by development teams during the product development process to aid in designing for calibration of trust. It consists of four parts: A sensitizing session package, autonomous function visualization canvas, user decision matrix and trust enhancing communication. Collectively, the four parts allow for addressing the two challenges as selected previously. Each part of the toolkit was tested with designers and engineers and further iterated. The complete toolkit was validated by conducting interviews with experts and triangulating the data with the test data gathered during the testing phase of the design process.

The testing and validation of the final outcome shows merit in the use of the toolkit for designing for calibration of trust and at the same time provides the flexibility for further modifications and adjustment. During the testing phase the participants found the use of the toolkit easy and intuitive. The digital method of testing suggested the deployment of the toolkit was possible in a digital setting. However, there were certain limitations to the project, the toolkit was not tested as a whole because of the time required and the unavailability of the necessary stakeholders. These limitations have been detailed out in the recommendation section of the report. Further research directions have also been suggested as a continuation of this project or start of new projects.

In conclusion, the project is a step in the right direction when designing for calibrated trust by building on the work of other researchers like Ekman et al.(2016) and Mirning et al.(2016), but requires further research and design in other areas to fully realize the idea of designing for calibration of trust, such as the work of Anika Boelhouwer at TU Twente and David Abbink at TU Delft . In a broader perspective the insights and toolkit designed should not be limited to autonomous vehicles but extrapolated to designing other social robots or autonomous technologies that will coexist in future societies.

Reading Guide

Welcome!

The reading guide is to provide some pre-requisite information to make the reading experience more enjoyable and less confusing.

The report is broken down into three parts:

- **Part-1 :** *Provides an overview of the project, literature review and the research method used for data gathering and analysis.*
- **Part-2** *: We begin part two with the reframed design question, explore the ideation process and the calibrated trust toolkit.*
- **Part-3***: The testing and validation of the toolkit along with further research, project limitations and personal reflection are subsets of this part.*

Within the project there are certain abbreviations that have been used. While these abbreviations have been described within the text, they have been collectively described in this section:

AV : Autonomous vehicle
DARPA : Defense Advanced Research Projects Agency
SAE : Society of Automotive Engineers
V2V : Vehicle to Vehicle
V2I : Vehicle to Infrastructure
NTSB : National Transportation Safety Board
VRU : Vulnerable Road Users
RU : Road Users
OEM : Original Equipment Manufacturers
SWOV : Institute for Road Safety Research, Netherlands
HMI : Human Machine Interface
eHMI : External Human Machine Interface
MaaS : Mobility as a Service
TU Delft : Technical University Delft
TU Eindhoven : Technical University Eindhoven
HuLAO : Human Centred Levels of Automation
TEC : Trust Enhancing Communication
TET : Trust Enhancing Technology
USP : Unique Selling Point

Some helpful definitions of certain terms/ phrases used within the project, that are not defined within the report:

Autonomous Vehicles (AV): An autonomous vehicle or a self driving car, refers to any vehicle that can sense its environment and operate without human involvement. There exist levels within an AV. But for simplicity the term autonomous vehicle represents all levels of automation within the report.

Human Machine Interface (HMI): It is a combination of software and hardware that present information to the user/ operator about the state of the process and to accept and implement the users instructions.

Pre-use phase: This represents the time from the end of the product development process of the vehicle to the instance at which the vehicle is sold to its owner.

Use phase: It is the subsequent phase to the pre-use phase and represents the time from when a vehicle is purchased by the user till the time the vehicle is discarded.

The report also consists of hyperlinks. The table of context is hyperlinked to the respective sections. Also, there are hyperlinks within the text which are represented by the text underlined with a blue line. Example : Design for Calibrated Trust

Reading in Digital Version:
The digital version of the report will open in a double page format. The ideal zoom for reading the report is 100%-115% for laptops screens.

"Trust, but Verify"

Russian Proverb

Photo by Luis J. from Pexels

Table of Content

Part 1 I Chapter 1

Project Context & Approach

This chapter provides an overview of autonomous vehicles, highlighting the benefits of a mobility system based on autonomous vehicles and the challenges we face in realising that future. We then explore the reason for selecting trust as the main focus of the project and create the scope of the project. The second part provides an overview of the design approach followed in the project

1.1 I Introduction

Introduction to Autonomous Vehicles

The conception of self driving vehicles or autonomous vehicles goes back nearly a century. The first effort to create an autonomous vehicle was a radio controlled driverless car in the 1920's (Davison & Spinolas, 2015). Since, then researchers and car manufacturers have tried to realize this dream of autonomous mobility. In 2004 DARPA (U.S Defence Advanced Research Projects Administration), launched a challenge to create an autonomous vehicle with the winner receiving $1 million (Davison & Spinolas, 2015). While the first year of the DARPA challenge did not bring much success in terms of delivering an autonomous vehicle, it did spark a revival of interest in the development of autonomous vehicles (Weber, 2014).

Fast forward to 2020, Zoox an autonomous taxi company founded in 2014 was acquired by Amazon for $1.2 billion(Rangwala, 2020). This is not the first deal to take place over the past decade with regard to autonomous technology. Nearly $1.5 billions have been spent by the most prominent car manufacturers to realize a fully autonomous vehicle. So the question here is why are companies spending large sums of money in the research and development of autonomous technology and what would be the effect of autonomous vehicles in the future of mobility? We answer this question in the next section and build on further to the scope of the project.

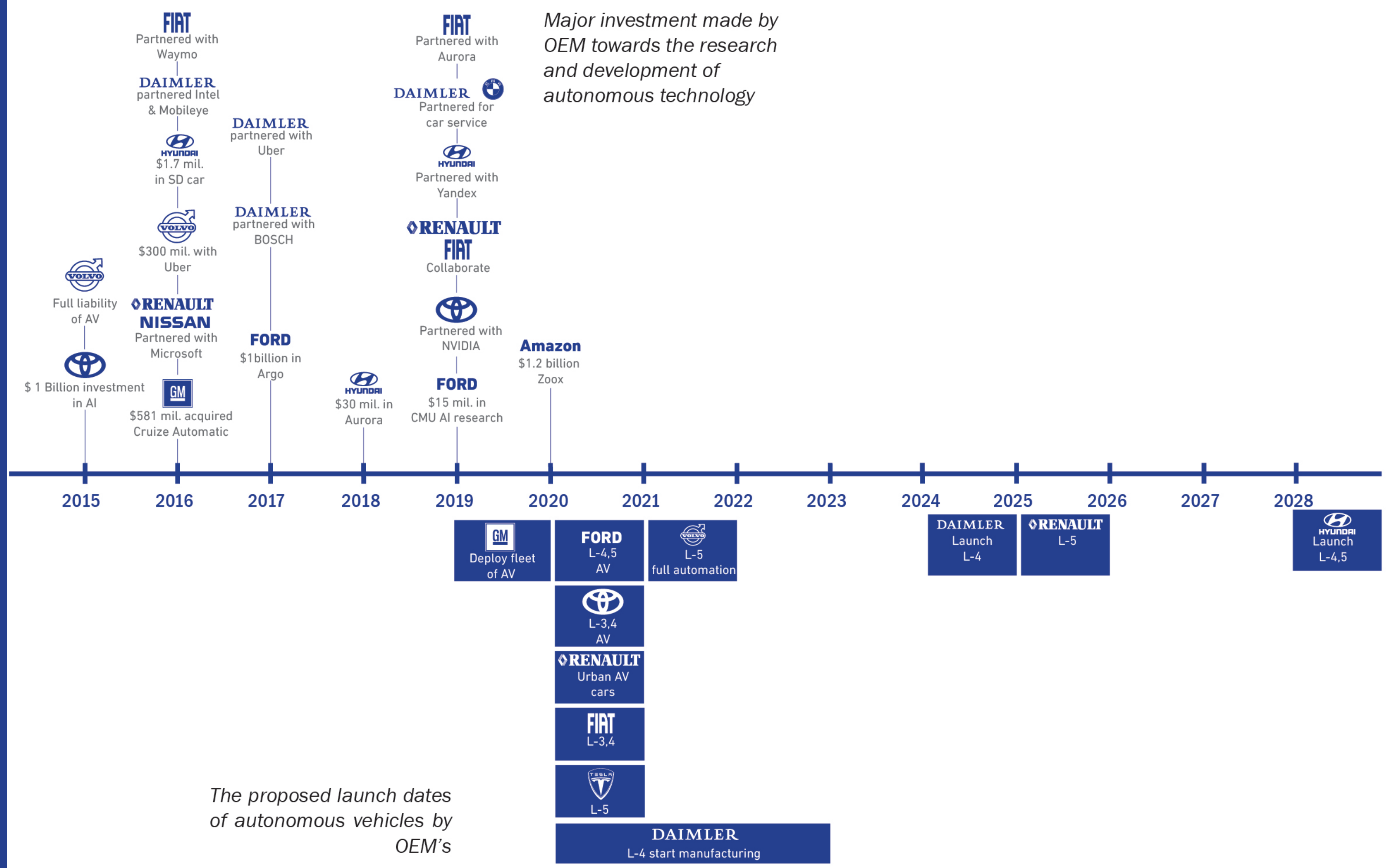


Figure 1.1. A timeline of investment and partnerships by OEM related to autonomous vehicles. The lower part of the timeline represents the proposed dates of launching autonomous vehicles and the level of automation.

Autonomous Vehicle Classification

Before we go further it is imperative to highlight the classification that is used to describe autonomous vehicles. This information will be directly used in the latter parts of the project and would be helpful to get an overview before moving further. The Society of Automotive Engineers (SAE) classifies AV into 6 categories, these are called levels and start from Level-0 and go up to Level-5. Level-0 is a vehicle which has no autonomous functionality and Level-5 represents vehicles with full automation. There exist other classifications of AV by various governing bodies, but for this project we will be referencing the classification as provided by SAE.

The Promise of Driverless Cars

The benefits of autonomous vehicles (AV) are more speculative than verified, yet the diversity and impact of the benefits makes the investment in them worthwhile. The most prominent benefit of AV is increased road safety. Nearly 90% of road accidents in the USA occur due to driver error and 40% of fatal accidents are caused due to the combination of alcohol, distraction, drug involvement and/or fatigue (Fagnant & Kockelman, 2015). The introduction of AV not only provides more safety but also is economically beneficial, according to the National Highway Traffic safety Administration report of 2012, nearly $227 billion is the annual economic cost of road crashes.

Aside for safety the second impact area is traffic operation, investigation into the influence of AV on traffic has shown that traffic congestion can be reduced. This could lead to increase in fuel economy by 23-39% (Fagnant & Kockelman, 2015). In addition to this the integration of vehicle to vehicle (V2V) and vehicle to infrastructure (V2I) communication have the potential to improve the lane effective capacity as well. In the long term the AV will make travelling more streamline and efficient.

From the perspective of the end consumer, AV provides the platform to put travel time to better use. Instead of waiting in traffic users can use this time to complete other work. This not only aids in the ability of multitasking but also makes it more safer for users (Fagnant & Kockelman, 2015). In addition to these benefits there are numerous other benefits of fully autonomous vehicles that have been proposed by researchers. This includes the introduction of mobility as a service (Maas), the use of ride sharing and in general an evolution of individual mobility (Wong *et al*., 2018).

While, these benefits of autonomous vehicles have been well documented. The fact remains that very little of this potential has been fully realized. On the contrary many barriers have seemed to emerge. In the next section we dive into the barriers to AV and take a step towards developing the scope of the project.

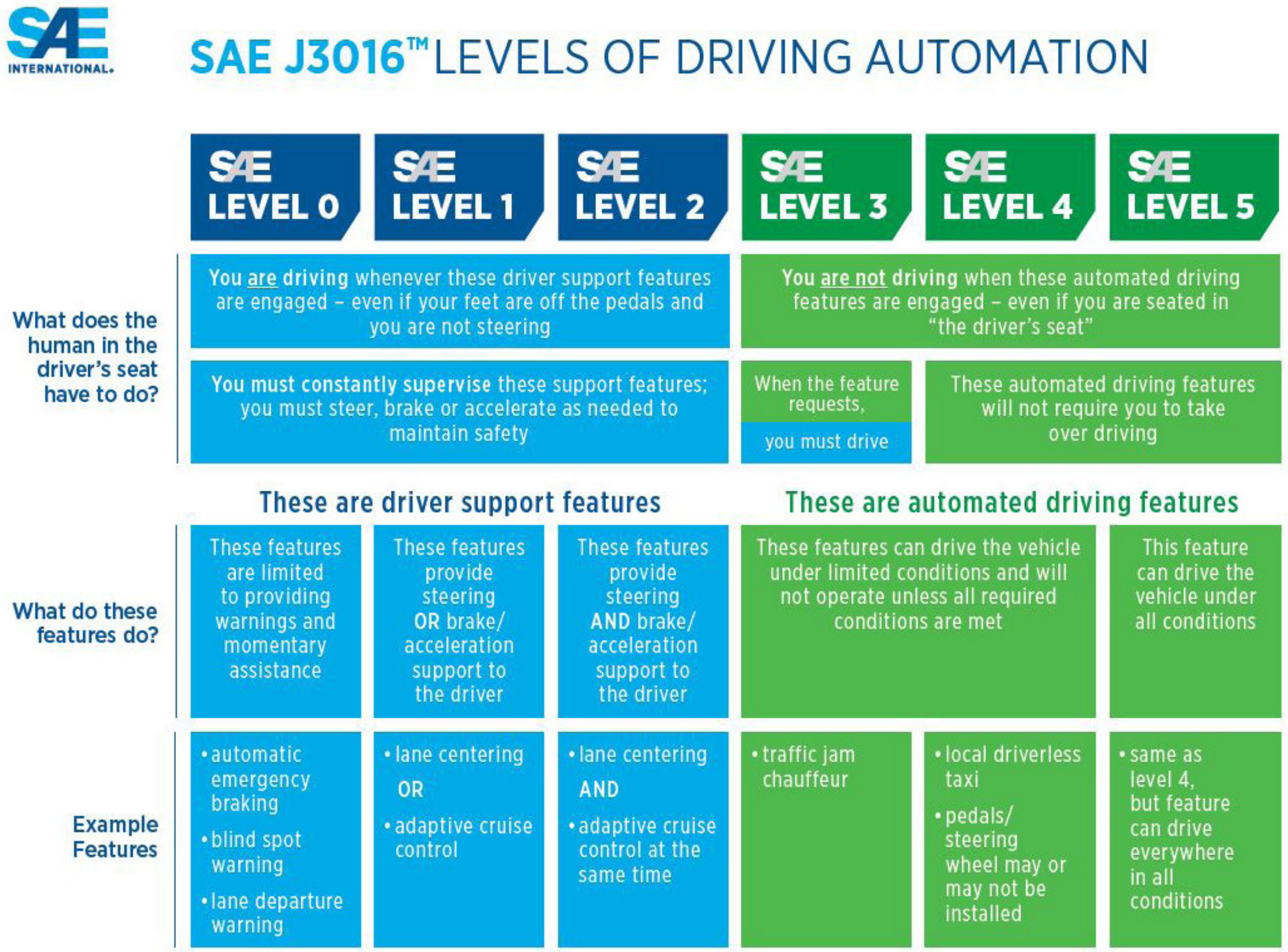


Figure 1.2. The SAE Internationals "Level of Driving Automation" according to the *SAE J3016: Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems*, 2018

The Barriers for Driverless Cars

The challenges faced in making AV a reality stem from a variety of different places, including but not limited to technological limitations (developing better AI models, collecting the adequate training data, cost of sensors and their reliability), ethical dilemmas (The most common one discussed is the trolley problem and variations of that problem), value tensions (privacy of users, autonomy of using AV) and human factor challenges (Misuse of AV). *Table 1.1* lists some of the challenges that have been highlighted in literature. The table by no means is exhaustive in nature and like how there exist a multitude of benefits of AV, there exist a mountain of challenges in the way to reaching these benefits.

The focus of the project is towards addressing two specific barriers: *Over reliance on automation* (Trimble, 2008) and *lack of trust in capabilities of AV* (Fraedrick and Lenz, 2014). The prominence of these barriers is illustrated by the recent accidents in Taiwan involving a Tesla operating on autopilot crashing into a toppled over truck. The reason for the accident was attributed to negligence of the driver (over reliance) and the inability of the AI to identify the toppled truck (Templeton, 2020).

Both of these barriers can be traced back to the lack of appropriate/calibrated trust and are the two polar opposites of the trust scale. In the current deployment of AV by OEM these two challenges have come to the forefront of the causes of accidents and thus making addressing them becomes imperative for the successful deployment of AV.

Barriers	Source
Over reliance on machines	Trimble (2008)
Loss of competencies and skills	Bazilinskyy *et al*.(2015)
Lack of trust in the capabilities of AV	Fraedrich and Lenz (2014)
Specific risks for crashes	Daziano *et al*.(2016)
Systems failures	Fagnant and Kockelman, (2015)
Breach of Information	Fagnant and Kockelman, (2015)
Deprived from joy of driving	Fagnant and Kockelman, (2015)
Electronic security concerns	Schoettle and Sivak (2014)

Table 1.1. Barriers identified within literature pertaining to the acceptance and deployment of autonomous vehicles

Why focus on Trust?

The project focuses on trust, more specifically calibrated trust between users and their autonomous vehicles. The reason for selecting trust as the main concept to design towards stems for the two barriers discussed in the previous section. As specified earlier the overreliance on machines and lack of capabilities of AV are the two extremes of the trust spectrum, with calibrated trust at the center (We will define calibrated trust in detail in the later section of the project).A second perspective for selecting trust lies in its importance towards the acceptance of autonomous vehicles (Adnan et al. 2018). Trust is considered as one of the most important factors that influence user acceptance and the adoption of autonomous driving into our society (Alawadhi et al., 2020). Moreover, we see that the formation of trust between a user and an autonomous vehicle is further influenced by other factors such as safety (under technology readiness), privacy (under legal readiness) etc. This not only makes the designing for trust important but also a pathway to explore the influence of other factors on the user acceptance of AV.

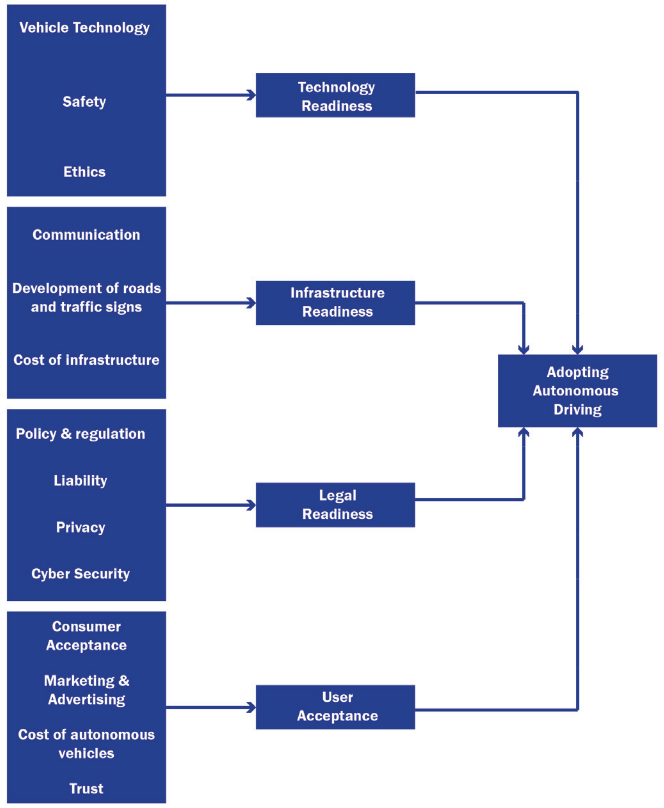


Figure 1.3. Autonomous driving factors conceptual model. Reprint from "A systematic literature review of the factors influencing the adoption of autonomous driving" by Alawadhi et al., 2018

Scope of the Project

The scope of the project is to design for calibrated trust, which is also known as appropriate trust. The reason for selecting calibrated trust as the scope in addition to the reasons described in the previous sections is because it not only benefits the user in using autonomous vehicles in a safer and better manner, but also aids the design and development team to understand the limitations and advantages of the autonomous vehicle (Lee & See, 2004). Thus, calibration of trust is not just user focused but also designer focused.

The Stakeholders

The project's main stakeholders are the autonomous technology development team at an automobile manufacturing company. These teams generally consist of the subsequent sub-functions: Hardware Engineers, Software Engineers, System Designers and Industrial Designer/ Engineer(It should be noted at this stage that depending on the company the team composition, size and the name can vary, however the core sub-functions as described will not change) . There are also secondary stakeholders, these are the other systems present within an automobile: body structure, chassis system, power-train system, fuel system, electrical system, climate control system, safety and security system and driver interface system. These stakeholders do not directly influence the development of autonomous technology but play a major role in integrating the autonomous technology into the overall vehicle. A more detailed overview of the stakeholders and their role in the product development process will be discussed in the latter stages of the report.

There is a second classification that is also used in the project for representing stakeholders. This divides stakeholders as internal or external stakeholders. Internal stakeholders are people who are employees of the company. External stakeholders represent all other people/users/organisations that are associated with the project.

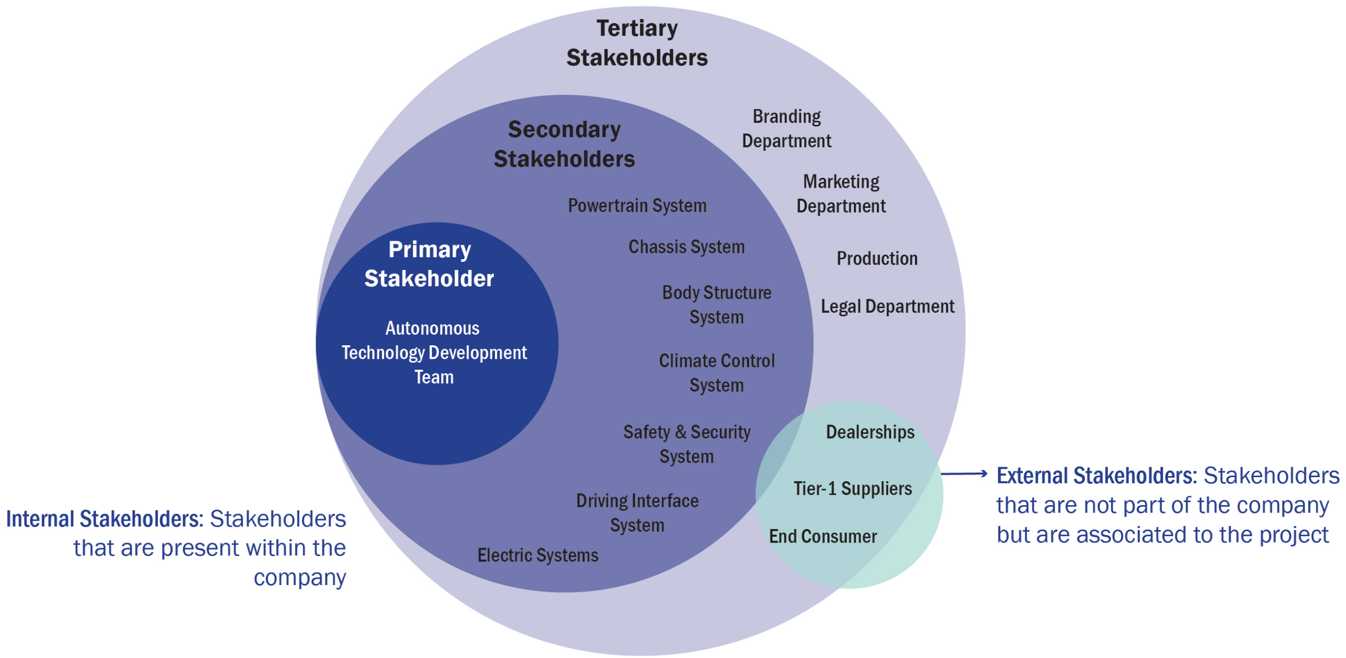


Figure 1.4. A visualization of the various stakeholders related to the development of an autonomous vehicle

1.2 | Design Question & Approach

Design Question

The design question selected to begin the project was :

"How to allow design teams of autonomous vehicles to design for human values (trust), for a multiple user context within urban locations ?"

The question was framed in this manner to allow for open exploration of the problem space and narrow down the reframed design question in the latter stages of the project.

Design Approach

The design approach for the project is based on the double diamond. The four major stages of the project are: *Discover*, *Define, Develop* and *Deliver* (Calabretta, Gemser & Karpen,2018). We begin with a literature study of the existing research in the area of calibrated trust, from then we move to selecting a research method to generate insights into the process. These steps complete the Discovery Stage. In the Define stage the analysis of the gathered data is performed and we end the first diamond with a reframed design question. The second diamond begins again with additional exploration within literature and the beginning of the ideation on the design question. Having selected the final design interventions they are evaluated with the stakeholders for improvements, this is the beginning of the last stage i.e. Deliver which involves creating the final design toolkit and validating it with experts.

While the four stages have clear boundaries within the double diamond the project is a lot more iterative and nature and thus there was a considerable amount of overlap between the stages especially within the first diamond. The detailed steps taken during the project can be found in the respective sections of the report.

Desired Outcome

The outcome of the project will be a design toolkit that can be used by the autonomous technology development team to design for calibrated trust of the end user. The reason for selecting a toolkit is based on the level of human experience the project targets. As the level is at an organisational level we see that using a design toolkit is an appropriate outcome (Friedman & Henry, 2019). In addition, designing a universal toolkit provides the opportunity to use a higher level of human experience such as social policy or global.

NOTE: *The complete project was conducted during the mandatory lockdown due to the COVID-19 pandemic. Due to which certain design decisions were taken to meet the rules set during the lockdown. While the overall process of the project is not affected greatly by the COVID-19, certain parts of the project were influenced. The reasons for the modification of these parts and the subsequent assumptions/actions have been discussed in their respective sessions in the report whenever necessary.*

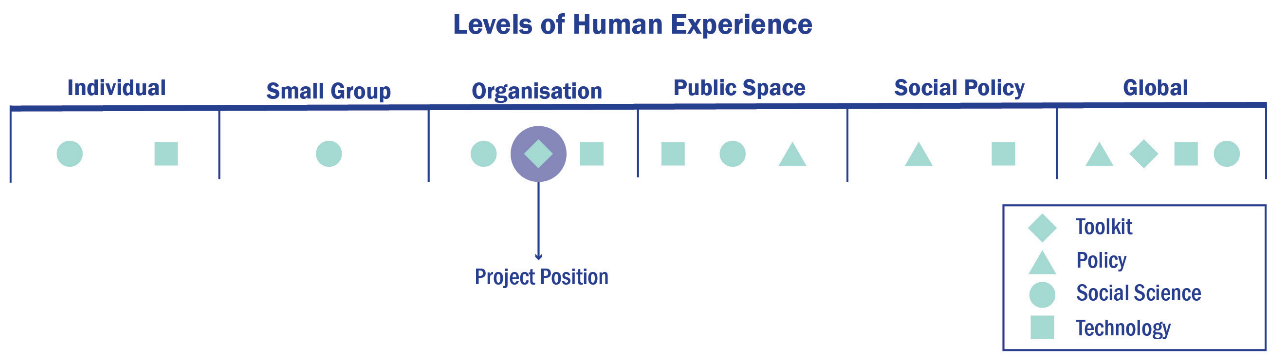


Figure 1.5. The levels of human experience, reprint from "Value Sensitive Design: Shaping Technology with Moral Imagination", Friedman & Henry (2019)

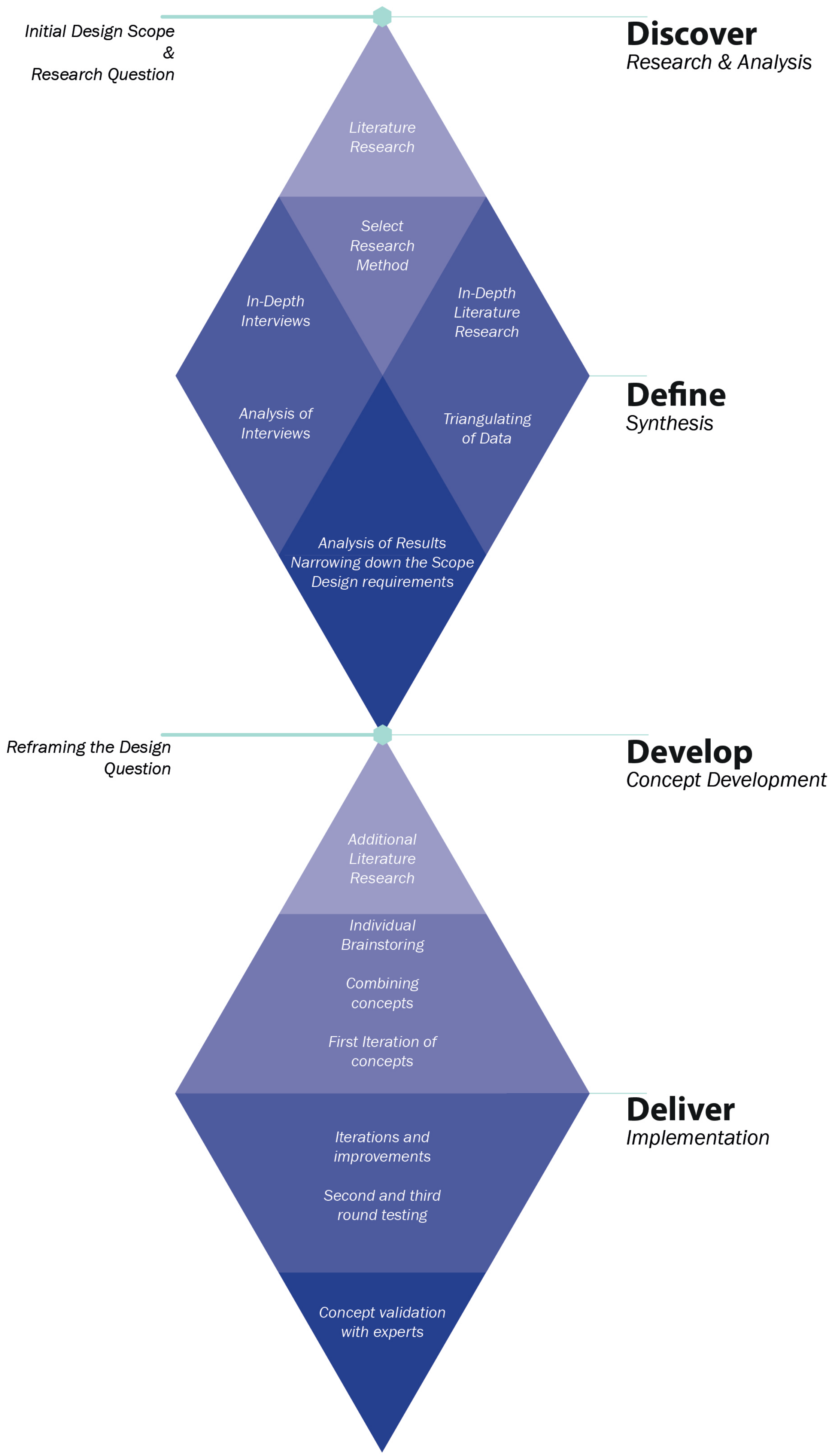


Figure 1.6. Representation of the double diamond design method used in the project

"Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less"

Marie Curie

Photo by Casey Horner on Unsplash

Part 1 I Chapter 2

Literature Review

The chapter covers the literature review section of the project. We begin with building an understanding of trust, formation of trust and the concept of calibrated trust. We then move onto the changing role of the driver, vulnerable road users and lastly human values and AV. The chapter is concluded by highlighting 10 key conclusions.

2.1 I Introduction

Overview

The literature review for the project spanned a number of different research areas. Some are in line with autonomous vehicles whereas others do not have a strong association with the research and development on the field of autonomous vehicles but are still relevant to the project. In addition literature research was also performed during the other design phases such as the ideation phase and validation phase. However, for ease of understanding all the literature research has been added into one section.

The initial questions that guided the literature research were:

- *What human values constitute the decision of placing trust in autonomous vehicles?*
- *How does trust work between technology and users?*
- *What role does trust play in the acceptance of autonomous vehicles?*
- *How to translate the above reasons into meaningful experiences for designers and development teams to create autonomous vehicles that users are willing to use?*

These questions were used to filter out the relevant research for the literature review. Following this process, the major themes that were identified to be researched were: *Trust between Human Automation*, *Role of trust in user acceptance, changing role of driver with introduction of autonomous vehicles, autonomous vehicles interaction with other road users* and lastly the *role of human values in the user acceptance of autonomous vehicles*.

The subsequent part of the literature review follows the same order, we start out with a brief understanding of trust and trust in human automation. We then focus on the concept of calibrated trust and its importance. Having created an understanding of trust we move on to focus on the various end users, the first being the driver itself. We explore the changing role of the driver and what are the benefits and challenges of driving an autonomous vehicle. We then look at other road users, especially vulnerable road users (cyclists and pedestrians). We conclude the literature with exploring the human values aspect of trust and acceptance of trust.

The above research was conducted to build the project, but as stated earlier secondary literature review was also conducted in subsequent stages of the project. This included research in the field of translating human values into design requirements, designing for socio-technical systems and the development process of automobiles.

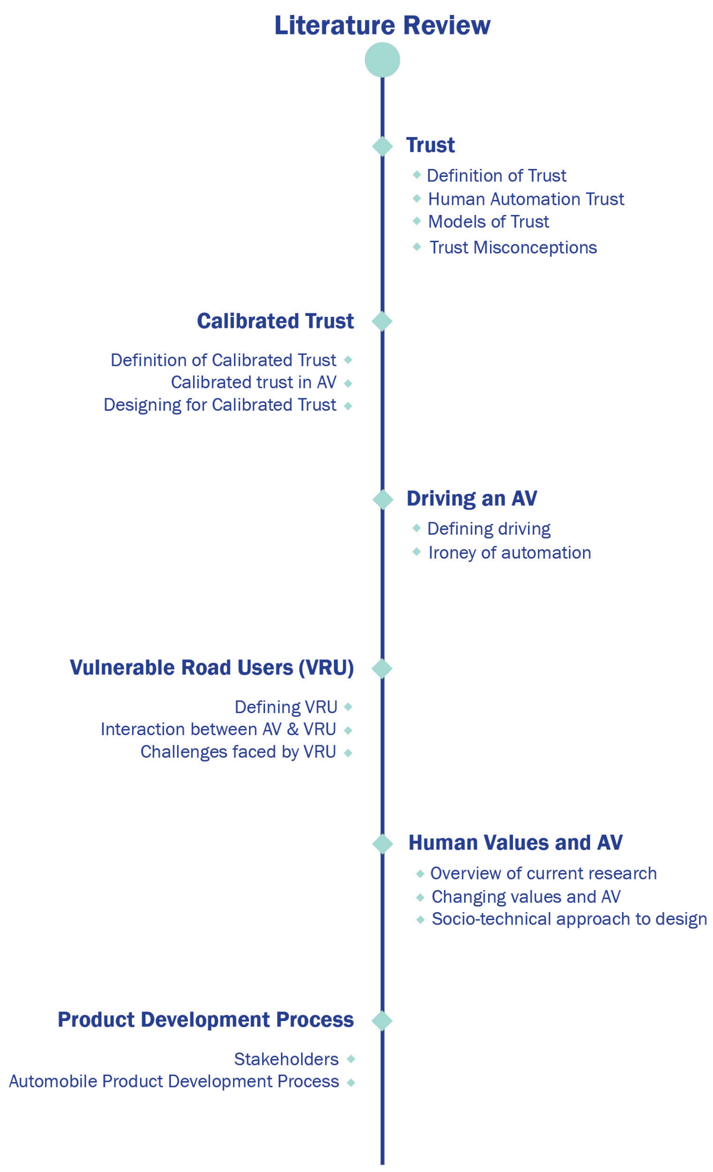


Figure 2.1. The various topics covered within the literature review of the project

2.2 | Trust

What is Trust?

The definition of trust has seen to vary from research to research depending on what they consider trust as (attitude/ intention/behavioural intent). From the literature we can see that trust can either be considered as an attitude, intention or even a behavioural result. Ajen and Fishbein (1980) developed a framework to overcome these conflicting definitions of trust (*Figure 2.2*). Through this framework we see that trust when considered as an attitude that affects reliance. Beliefs underline trust and various intentions and behaviours may result from the different levels of trust (Lee & See, 2004). From this framework Lee & See (2004) came up with a simple definition, which we will be using as the working definition of trust within the project as well: 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"

The definition of trust by Lee & See (2004) does provide a starting point for diving deeper into the formation of trust. Mayer et al.(1995) concluded that there are three levels that form the basis of trust: ability, integrity and benevolence.

Ability: It consists of the group of skills, competencies and characteristics that enable the trustee to influence the domain.

Integrity: The degree to which the trustee adheres to a set of principles that the trustor finds acceptable

Benevolence: The extent to which the motivation and intent of the trustee are aligned with those of the trustor (Lee & See, 2004).

To understand the basis of trust as highlighted earlier, we consider interpersonal trust first and then move onto the idea of trust in automation. Interpersonal trust is considered an evolving phenomena with trust constantly changing (Rempel et al., 1985). They proposed that trust is a factor of predictability, dependability and faith. Predictability, is the degree to which future behavior can be anticipated(this is similar to the idea of ability). Dependability, is the consistency of the behavior (equivalent of integrity). Faith is a more general judgement that a person can be relied upon (equivalent of benevolence) (Lee & See, 2004)

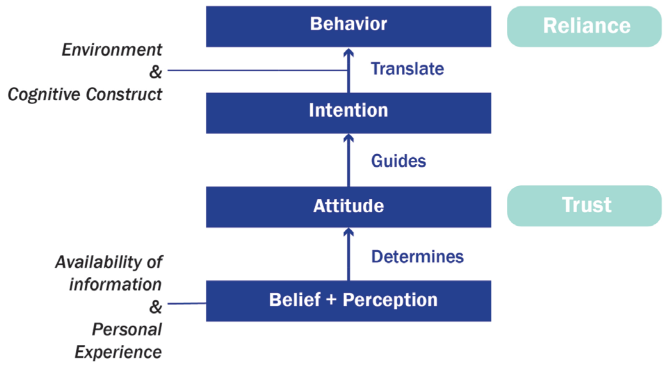


Figure 2.2. The visual representation of the framework as proposed by Ajen and Fishbein (1980)

Human Automation Trust

A similar distinction was created by Lee & Moray (1992) for human automation trust, which has been described as follows:

Performance: It refers to the competency or expertise an automation has to achieve a specific goal. Performance includes characteristics such as reliability, predictability and ability of the automation.

Process: It is the degree to which the automations algorithm are appropriate for the situation and able to achieve the operator's goal (Lee & See, 2004). It is the ability of the user to understand the algorithm and seem capable of achieving the operator goals in the current situation.

Purpose: Refers to the degree to which the automation is being used within the realm of the designers intent (Lee & See, 2004).

While the above discussion shows that there are many similarities between the notion of interpersonal trust and human automation trust. There are some fundamental differences, the most prominent one being lack of intentionality. This arises from automation not having their own purpose, which is not the case in the scenario of interpersonal trust. However, automation might not have their own intentionality; they will embody the intentionality of the designer (Rasmussen, Pejterson & Goodstein, 1994).

Type of Trust	Source	Trust Formation Process		
Trust in General	Mayer et al.(1995)	Ability	Integrity	Benevolence
Interpersonal Trust	Rempel et al.(1985)	Predictability	Dependability	Faith
Human Automation Trust	Lee & Moray (1992)	Performance	Process	Purpose

Table 2.1. The representation the various trust formation process as proposed in the literature

Another interesting difference between interpersonal trust and human automation trust lies in the evolution of trust. In the case of interpersonal trust, we start with predictability, which progresses to dependability and evolves to faith. Human automation trust can also follow this path but is capable of following the opposite pattern, i.e. start with faith then dependability and finally predictability (Lee & See, 2004).

In conclusion, while there exist stages in which trust is developed, it is not necessary that those stages have to be followed. Trust development primarily is dependent on the information available rather than a fixed series of stages (Lee & See, 2004).

Model of Trust

The discussion with regard to trust till now has been focused on its definition and the formation of trust in novel situations. However, trust is a complex attitude that is influenced by not only the truster and trustee but the context as well. On of the prominent models of trust is by Lee & See (2004), for this project we will be using the model proposed by Hoff & Basihir (2015), which build on the initial mode of Lee & See (2004) and provides a more detailed overview of factors influencing trust as seen in *Figure 2.3*. Hoff & Bashir (2015) divide trust into three parts: *Dispositional trust, Situational Trust and Learnt trust*. In addition to the three parts of trust, the model also highlights which part of the trust is influenced prior to interaction and during interaction.

Dispositional Trust: represents an individual's overall tendency to trust automation, independent of context or a specific system (Hoff & Bashir, 2015). Factors such as culture, age, gender and personality traits influence is considered in this type of trust.

Situational Trust: It represents the context in which the automation is used. Situational trust can be further broken down into external variability, that includes variables such as workload, task difficulty, type of system and the system complexity. Internal variability which includes factors such as self-confidence, mood and attention. This is more focused on the current state of the user when interacting with the automation

Learned Trust: The final part of trust development is learnt trust which is a representation of a users evaluation of a current system drawn from the past experience or the current situation (Hoff & Bashir, 2015). Learned trust is further divided into two sub parts: *Initial learned trust* and *Dynamic learned trust*.

In addition to the three types of trust there are other situational factors that influence trust along with design features of the automation. From the model presented by Hoff and Bashir (2015), we see that there is a difference between prior to interaction and during interaction trust. It is worth noting that during interaction trust can change at every instant based on how the context of use changes. This is also highlighted by Hoffman (2017), "*the act of trusting is a continuous process of active exploration and evaluation of trustworthiness and reliability, within the envelope of the ever-changing work and the ever-changing system*". In addition we can also infer from the model that design features have a direct influence on the learnt trust part of trust. In terms of an autonomous vehicle this refers to the fact that the HMI system within the vehicle can only influence the learnt trust and the situational trust and dispositional trust will not be affected.

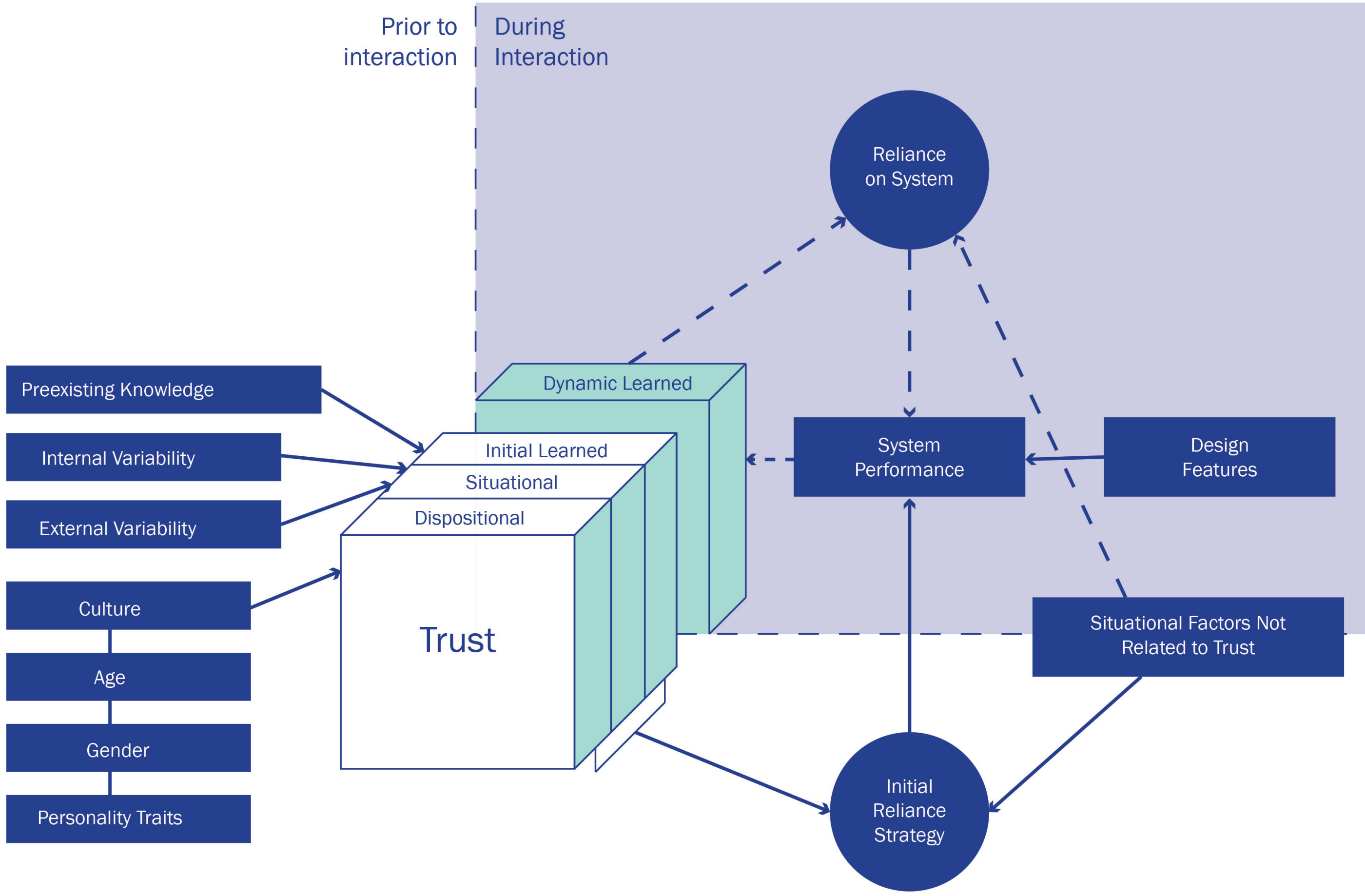


Figure 2.3. The full model of factors that influence trust. Reprint from "Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust", by Hoff & Bashir (2015)

Trust Misconceptions

It is important to highlight that at this point of time that there exist certain misconceptions (Hoffman, 2017) with respect to trust that need to be kept in amind while moving ahead. These misconceptions pertain not only to the theoretical knowledge about trust that we have but also the practical implications it can have. These misconceptions are:

Trust is measured on a single scale: There is no measurement method that can measure trust accurately using a single scale/ value. The measurement of trust is a collection of different factors and measuring one factor only provides quantification of that aspect of trust.

There is a point on that scale that serves as a metric; that is, trust is sufficient or falls at the "right" level: There is no right or wrong level of trust, nor is there a minimum threshold of trust. This builds on the previous misconception that we can quantify trust and once quantified can associate meaning to the values by creating specific thresholds.

When trust reaches that metric point, we want it to stay there. We want the trust to not change once it calibrates: Trust is ever changing and will always vary from one instance to another. Thus, there is no such point/threshold at which trust reaches equilibrium and does not change. Unlike mechanical instruments which have a standard calibration point, the calibration of trust is ever changing.

2.3 | Calibrated Trust

What is Calibrated Trust?

Having a working definition of trust and understanding the formation of trust between human and automation, we now move to the concept of calibrated trust. Calibration refers to the *"correspondence between a person's trust in the automation and the automations capabilities"* (Lee & See, 2004).

To understand calibrated trust it is important to perceive what would happen if there were no calibration of trust. From *Figure 2.4* we see that there exist two cases, the first is overtrust/mistrust and the second is distrust.

Overtrust/Mistrust: In this phase the trust exceeds the system capabilities leading to misuse of the system. Example of mistrust is the crash of Airbus A320 in 1995, the pilot trusted the ability of the autopilot and failed to intervene and take manual control. A similar case was seen with the Royal Majestic cruise ship that drifted off the course for 24 hours because the crew trusted the navigation system and failed to intervene (Lee & See, 2004). Similar incidents have occured with autonomous vehicles as well, in 2016 a Tesla in auto-pilot crashed into a truck crossing the highway. The cause of the accident was ruled as over-reliance on automation by the National Transportation Safety Board (NTSB) (Endsley, 2019)

Distrust: It occurs when the system capabilities exceed the trust users place on the system. Examples of distrust can be seen in the Costa Concordia cruise ship disaster that killed 32 passengers in 2012. The reason was because the captain distrusted the ships navigation system in favour of manual control resulting in the ship crashing into a shallow reef (Lee & See, 2004).

The examples highlighted in the discussion of mistrust and distrust are but a few prominent accidents that caught the attention of the world media. However, mistrust/distrust towards a system is a very common occurrence and can be seen in our day to day life. One such example is the use of applications on our smartphone. When we download and use an application we are directed to dialogue boxes that ask for confirmation if the application can access our data and hardware, such as camera, contacts, photos, etc. But how many times do we actually stop to read what we are giving permission for? This is also one form of mistrust. So, if it is a common occurrence for users to mistrust or distrust technology in a day to day life why is this an important topic when designing for AV?

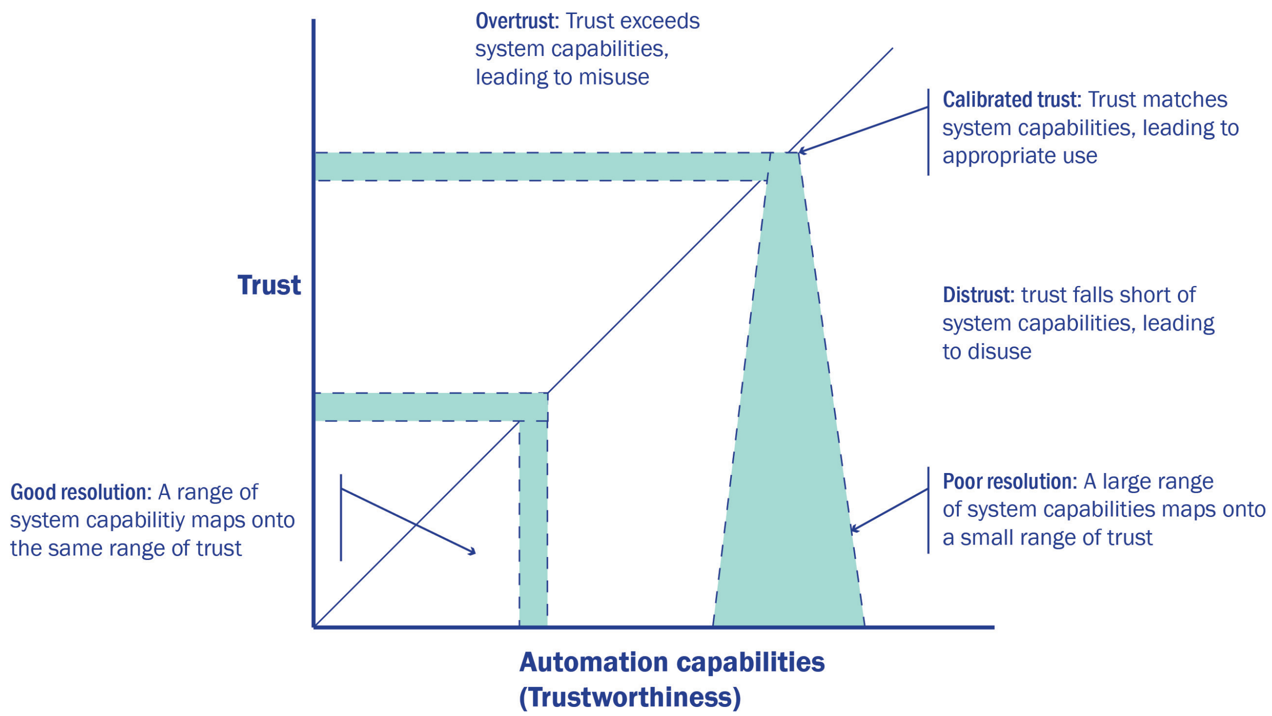


Figure 2.4. The representation of calibrated trust. Reprint from "Trust in automation: Designing for appropriate reliance", by Lee & See (2004)

Calibrated trust for AV

From the incidents discussed in the previous section, it can be seen that some of the most devastating accidents due to mistrust and distrust have occurred in the aviation and cruise lines industry. This is because automated systems were introduced in these fields much earlier than road vehicles. Further, there has been considerable research with regard to trust in these fields as well. However, the implications of mistrust or distrust can be higher on road user down the line due to two reasons:

Time to react: Reaction time to avoid an accident on roads is very short (in seconds or in some cases even less) as compared to the reaction time a pilot or a captain of an aeroplane or ship have to make adjustments to avoid a fatal accident.

Training: Pilots and captains have to undergo a lot of training and adhere to strict safety protocols, in comparison to driving which requires the passing of a single drivers test.

While the above reasons pertain to the user, the technology front is not very promising as well. The state of California has imposed policies for testing of autonomous vehicles to provide data to the state. From the data (till 2017) it is seen that Waymo (Google) can travel approximately 5,600 miles without a human driver having to intervene or before the vehicle disengages on its own due to some detected problem (Endsley, 2019). When comparing this to humans driving it takes on an average 490,000 miles between accidents and nearly 95 million miles between fatal accidents (Endsley, 2019). It should be noted that the disengagement of an AV does not correlate to an accident but that there was a failure(software, hardware, weather condition and road surface conditions). Thus,the figures are not a concrete representation of the problems with autonomous vehicles but highlights the need to design autonomous systems in a more reliable and resilient manner. This is also collaborated by the research of Kalra & Paddock (2016), who used statistical methods to find the distance an AV will need to travel before its rate of accidents better the current numbers.

Manufacturer	2015	2016	2017
Waymo	1,244	5,128	5,596
Nissan	14	146	208
Delphi	41	17.5	22.4
Mercedes-Benz	1.5	2	4.5

Table 2.3. The number of disengagements for autonomous vehicles. Reprint from "Situation Awareness in Future Autonomous Vehicles: Beware of the Unexpected", by Endsley (2019)

Can we design for Calibrated Trust?

Till now we have discussed what is calibrated trust and in theory it makes sense. But the term calibration in its conventional sense is the process of checking a measuring instrument to see if it is accurate. This raises the question: *What would be the instrument in this scenario of calibrating trust between the user and the autonomous vehicle?*

The answer to that question is the mental model of the user. However, this seems ironic as we are comparing a human and his mental model to a mechanical measuring device. But this is not the case, when we refer to calibration in terms of human automation trust we are referring to the dynamic information exchange between the automation and humans concerning the state of the automation that leads to reliance (Hoffman, 2017). This is the meaning of calibrated trust and the project also follows the same understanding of calibrated trust. Thus, calibration of trust is the active management of the users mental model when using an autonomous vehicle.

A mental model is created by a user in three steps: *understandability, predictability* and *reliability* Understandability represents the understanding of the capabilities of the autonomous system. Predictability refers to the ability to predict the actions of the autonomous system based on the understanding system. The last step is reliability, which represents the number of times the action can be repeated by the autonomous system without deviation from the ideal outcome. Calibration of trust can now be defined as the active management of the three aspects of the users mental model: *understandability*, *predictability* and *reliability* (Hoffman, 2017).

Designing of Calibrating Trust

Designing for calibrated trust is a relatively new topic, there is thus not much research on methods to calibrating trust. One of the first methods proposed was by Visser et al.(2014), their research is on creating trust cues for calibration trust. This is achieved by creating a matrix between the trust evidence level (origin, expressiveness, process, performance and intent) and the information process stage (perception, comprehension, projection, decision and execution). A similar framework for understanding calibrated trust was provided by Mirnig et al.(2016), The difference between their framework and Visser et al.(2014) was the use of task breakdown (operational, tactical and strategic) instead of trust levels.

Ekman, Johnsson & Sochor (2017) (*Figure 2.5*), considered the overall journey of using an automated vehicle, from entering the vehicle to exiting it. Highlighting 13 unique interactions in the overall journey and plotting them against 11 factors that can be used to calibrate trust at the required moment. Okamura & Yamada (2018) presented a different approach towards addressing the challenge. They proposed

the use of intuitive and logic on one axis and machine and anthropomorphism on the other (*Figure 2.4*).

The above discussed four research look at calibration of trust from different angles and provide a strong base to move forward from. However, there exist certain limitations as well. Nearly, all four have focused towards the use phase, which translates to addressing the dynamic learnt trust. There is still a gap in understanding how the pre-use phase can be used effectively in calibrating trust.

From the trust model presented by Hoff & Bashir (2015), we see that there are a number of factors that influence trust. So, does designing for calibrating trust mean addressing all of those factors? Hoffman (2017) provides an answer to this question by highlighting that trust is based on trustworthiness and reliability. Which require two criterion to be achieved: feedback and observability/understandability (Hoffman, 2017). Thus, *calibration of trust is the continuous process of balancing feedback and observability/understandability to attain an appropriate procedure to actively manage reliance* (Hoffman, 2017).

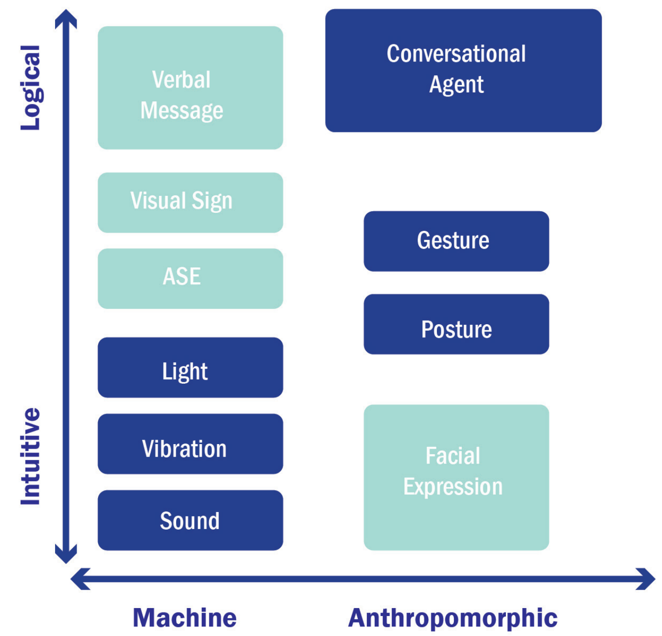


Figure 2.4. The representation possible design cues for calibrated trust. Reprint from "Adaptive Trust Calibration for supervised autonomous vehicles", by Okamura & Yamada (2018)

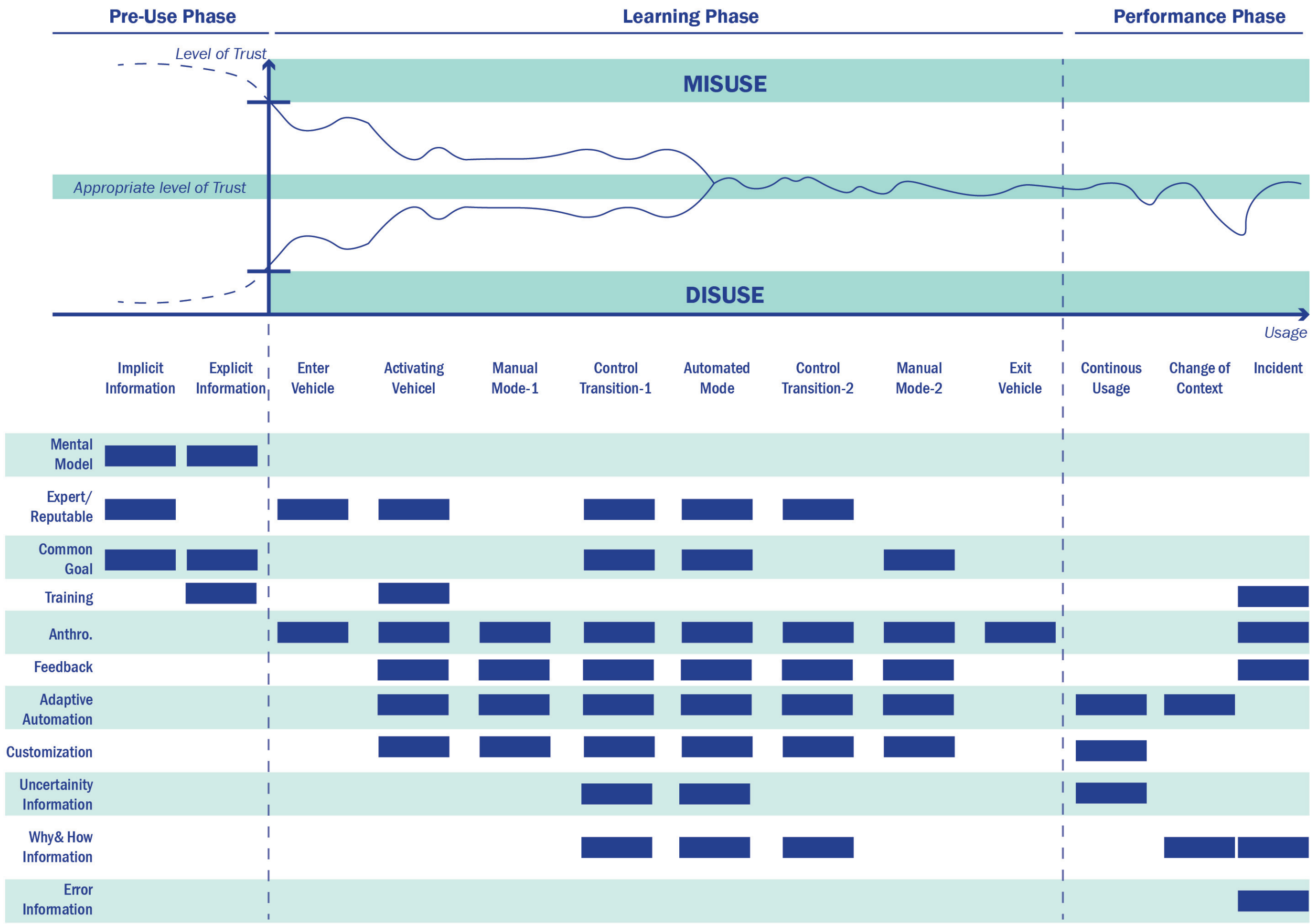


Figure 2.5. The representation possible design cues for calibrated trust. Reprint from "Creating appropriate trust in automated vehicle systems: A framework for HMI design", by Ekman, F., Johansson, M., & Sochor, J. (2017).

2.4 | Driving an Autonomous Vehicle

Defining Driving

The discussion till now has been focused towards a more general understanding of trust and calibrated trust. In this section we dive into the changing role of driving with the introduction of autonomous systems and how this influences trust. To understand how automation has influenced driving, we first need an understanding of the basics of driving.

The general theory of driving behavior was proposed by Fuller (2004), who stated that driving is the balance of capabilities and task demand. When the capabilities are higher than the task demands the driver will have control of the vehicle and as we move towards the opposite side of the spectrum where the demand of driving is higher than the capabilities there is loss of control which can potentially lead to collision of the vehicle. The introduction of autonomous systems within vehicles was seen as a method to substitute the capabilities of the driver or provide additional capabilities to match the task demand. While this makes sense in theory, in practice the results are slightly different. Banks et al (2017) analysed the task demand with and without automated systems for drivers and found that there was an increase in the task demands in autonomous vehicles when compared to normal vehicles (*Figure 2.7*). A similar outcome was seen when considering non driver related information and tasks, as we move towards higher levels of automation there is an increase in non-driver related information and tasks (Wang et al., 2018)(*Figure 2.8*).

These research results provide a glimpse into the future of driving autonomous vehicles and the fact that there is bound to be an increase in information and task demands placed on the user. However, going back to the general driving theory, we can see that there is also an increase in capabilities. Thus, the introduction of autonomous systems increases the capabilities of the driver at the same time the task demands related to driving. Do these two balance each out? is difficult to predict theoretically and would require simulation or naturalistic studies. This scenario does not hold true for all levels of automation. As we move towards full automation (level 4 and 5), the autonomous system will take nearly full control of the vehicle and thus the task demand for the driver will be minimal or null. The criticality of this balance is important in the semi- autonomous levels (level 2 and 3).

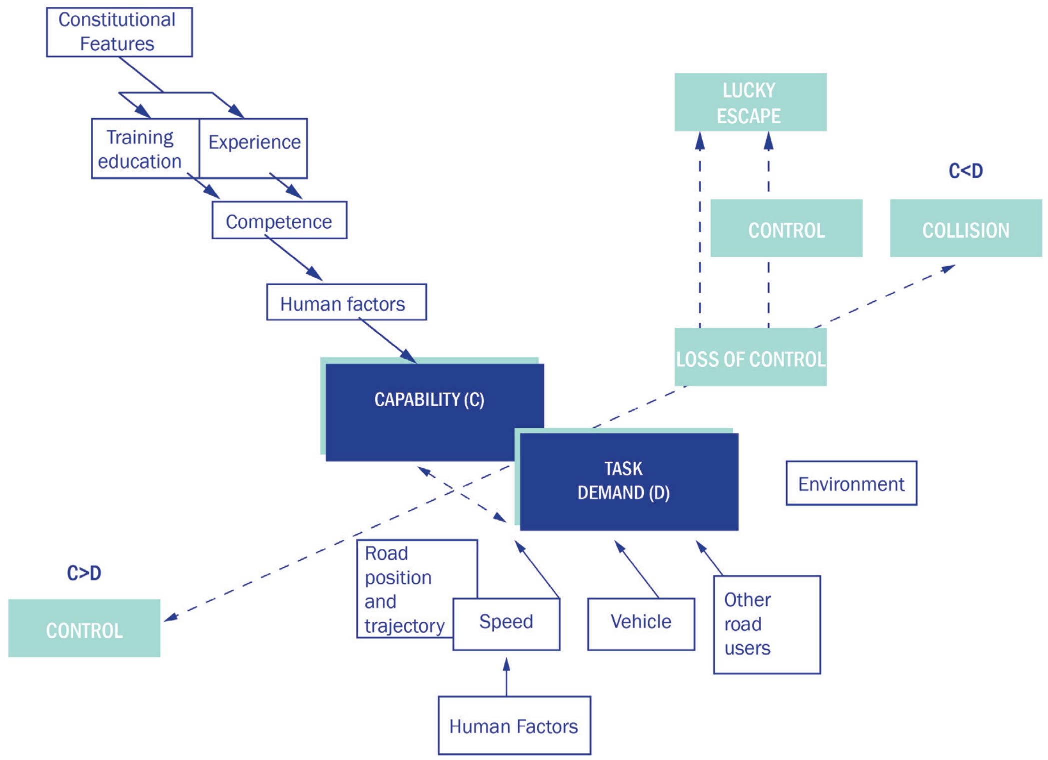


Figure 2.6. The representation of the task-capability interface model. Reprint from "Towards a general theory of driving behaviour"", by Fuller (2004)

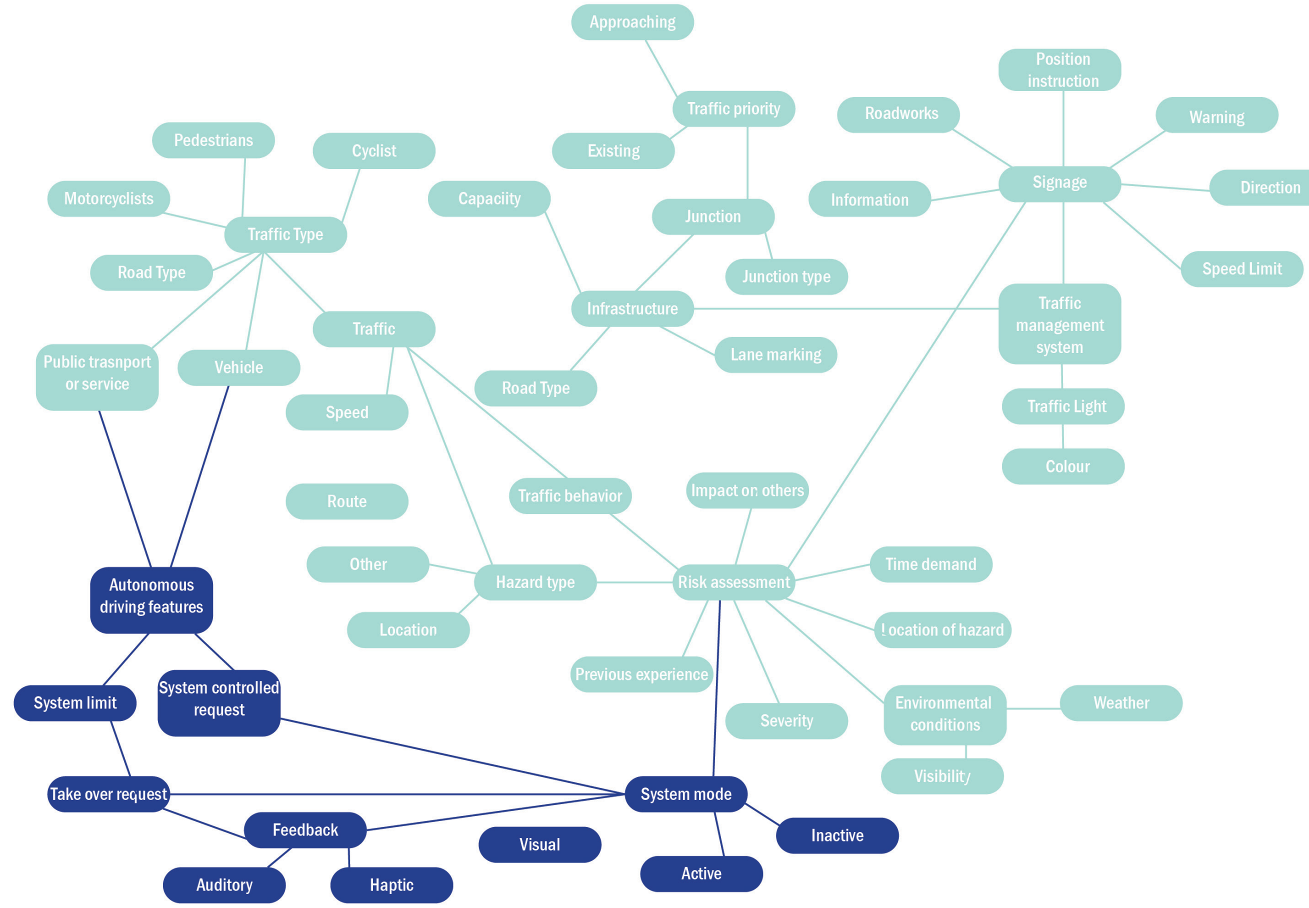


Figure 2.7. A schematic overview of the information a driver receives during the monitoring of an autonomous vehicle. The parts in dark blue are the additional information the driver needs to process as compared to conventional driving. Reprint from "Analysis of driving role: modelling the changing role of the driver in automated driving systems using EAST", by Banks et al (2017)

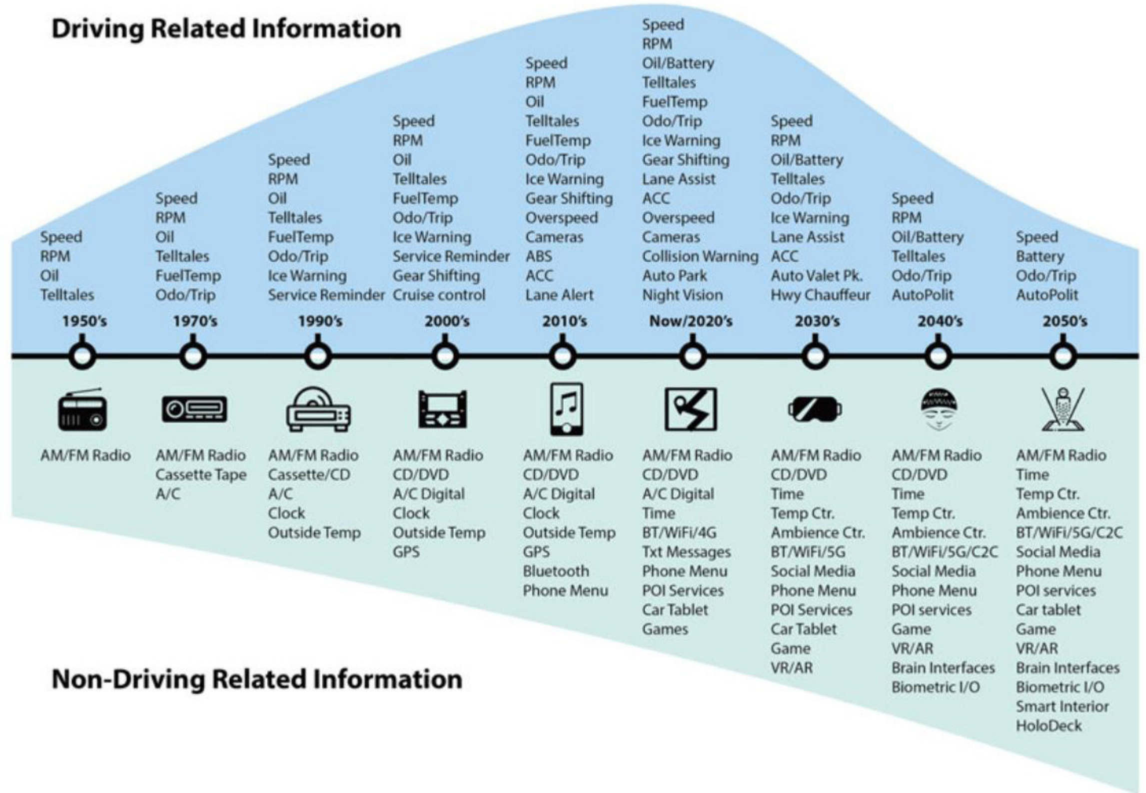


Figure 2.8. A schematic overview of the information needs changing in cars over time. Reprint from "Designing the Product-Service System for Autonomous Vehicles", by Wang et al (2018)

Irony of Automation

An autonomous vehicle is a form of automation and while we discussed the changing role of a driver in the previous section. We now focus on certain challenges that pertain to automation in general. More specifically we focus of three ironies that have been associated the autonomous vehicles form the very beginning:

Ironey 1: The higher the level of automation becomes, the more people will trust the automation and are likely to fail in taking over manual control during critical situations (Endsley, 2019). This has been labelled as the "lumberjack effect".

Ironey 2: When the workload is the highest, it offers the least assistance (Bainbridge, 1983). This means that in cases when there is the most need for automation are the situations in which automation seems to have the least influence.

Ironey 3: New role of users in automation is monitoring the system for abnormalities and alarms, whereas research shows that humans lose attention and psychological awareness when asked to monitor (Bainbridge, 1983).

The reason for calling these statements ironies is because in these cases we expect automation to aid the user but the opposite case exists. Interestingly, the three ironies selected are part of the 10 ironies of Bainbridge (1983) and nearly 40 years are still relevant.

In the last section of understanding the role of driving autonomous vehicles, we explore how drivers learn about autonomous vehicles technologies. Naturalistic studies and interviews with autonomous vehicle owners have provided insight into this question. One of the most prominent challenges faced by owners of autonomous vehicles is the lack of training and information provided by dealerships, owners manuals (Lin, Ma & Zhang, 2018). In addition it has been observed that 63% of owners currently learn to use the autonomous function through vehicle manuals and 59% by trial and error (Abraham et al.,2017) (*Figure 2.9*). On the contrary we see that ideally only 25% of the users would prefer to learn by trial and error, similar results were found by Boelhouwer et al.,(2020) within the Netherlands.

From the discussion we see that we are not only facing technological limitations of automation that makes driving a challenge. The method of teaching users is also not appropriate. However, research is being conducted in new ways of teaching and training drivers about automation focused on the "Car teaches me" category, under the I-Cave project, a joint venture by TU Eindhoven and TU Twente. Within TU Delft the VIDI:Symbiotic Project looks to find a new way of driving an autonomous vehicle by focusing on a symbiotic relationship between the autonomous system and the driver.

The relationship between calibrated trust and the evolution of driving an autonomous vehicle is aptly represented by the following statement by Casner et al., (2016) "*One of the most daunting challenges will happen when we reach the crossover point where automation systems are not yet robust and reliable enough to operate without humans standing by to take over but yet are too complex for people to comprehend and intervene in a meaningful way*". This statement relates to the section of "Can we Design for Calibrated Trust?". We see that the statement covers the factors of understandability, predictability and reliability of autonomous vehicles which make up the aspect of calibrated trust.

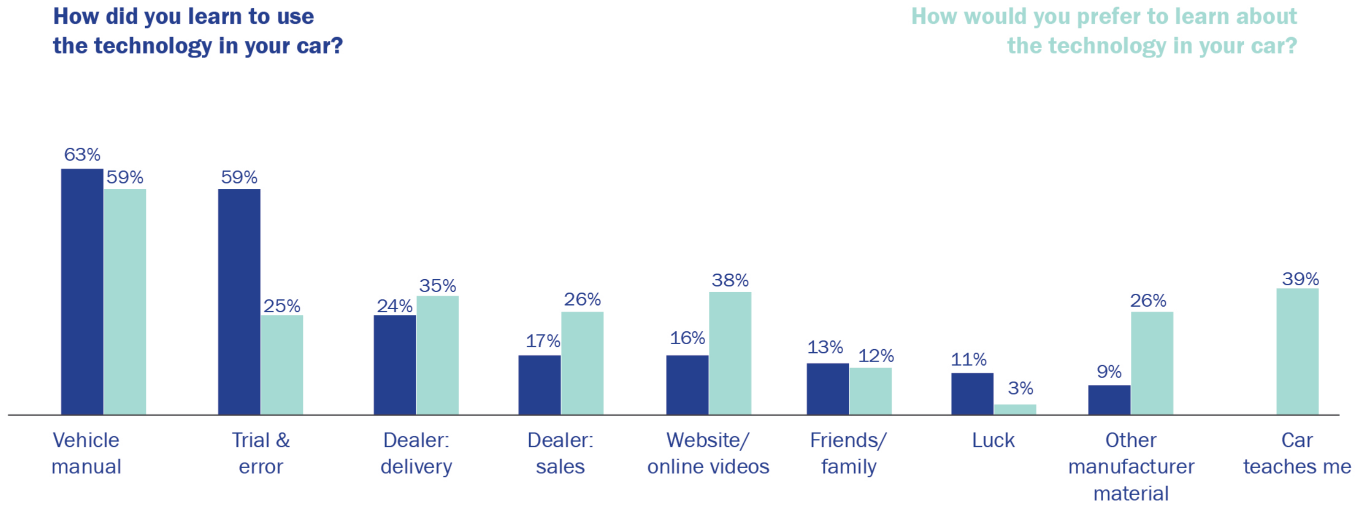


Figure 2.9. A schematic representation of the preferred method of learning about AV as compared to the actual scenario. Reprint from "Autonomous vehicles and alternatives to driving: trust,preferences and effect of age", by Abraham et al (2017)

2.5 | Vulnerable Road Users (VRU)

Defining VRU

In the previous section we looked at the influence of autonomous vehicles from the perspective of the driver, however there exist other road users on the road. In this section we focus on one specific category of road users i.e. vulnerable road users. The reason for selecting vulnerable road users was because they stand to gain a lot from the introduction of AV, but also can lose even more if an AV was to malfunction.

Vulnerable road users (VRU) can be defined in a number of different ways. They can be classified based on the amount of protection they have,or the amount of task capability. Institute for Road Safety and Research, Netherlands (SWOV) classifies vulnerable road users into the following categories: pedestrians, cyclists, moped and light moped riders, motorcyclists, children and elderly in traffic. There is a distinction made between pedestrians, children and elderly because children and elderly are subsets of pedestrians they have additional requirements, thus acting as a special case of pedestrians (SWOV Fact Sheet, 2012).

The classification of VRU can vary from place to place depending on the definition that is considered as seen in the case of SWOV. In this project we will use the classification based on the amount of protection they have. In addition the criteria of "*vulnerable should not themselves be a threat to others*", will also be considered. This is to remove novice car drivers and elderly car drivers. From this classification we can narrow down the initial list to: pedestrians, cyclists, moped and light moped riders.

Interaction between VRU & AV

The complexity of the interaction between autonomous vehicles and vulnerable road users is captured accurately by Rasuli & Tsotsos (2019) "*Driving is not just a dynamic control task but also a social phenomena and requires interaction between all round users involved to ensure the flow of traffic and to guarantee the safety of others....The challenge is to treat the problem not as a rigid dynamics problem but that of a social being*".

In a simpler sense the interaction of a vehicle and a VRU is not that of just following traffic rules, but is a social interaction in which the vehicle is also a participant. Further, the interaction is two way, between the VRU and the vehicle and vice versa. Research into VRU has shown that communication plays a major role in acceptance of AV. Habibovic et al.(2018) highlighted more than 20 communication cues that are used in daily life to communicate between vehicles and VRU. More importantly, 97% of the communication between vehicles and VRU takes place through implicit cues (Rasuli & Tsotsos, 2019). Implicit cues are defined as cues that do not directly provide the information required, but when combined with the information available in the current context provide the required information to make a decision. An example of this is eye contact between the driver and pedestrians. Thus, the first challenge to building trust within VRU is to create effective communication between the AV and the VRU, once this has been achieved we can move onto the idea of calibrating trust. However, communication is not the only challenge there are other prominent challenges that are also present which have been discussed subsequently.

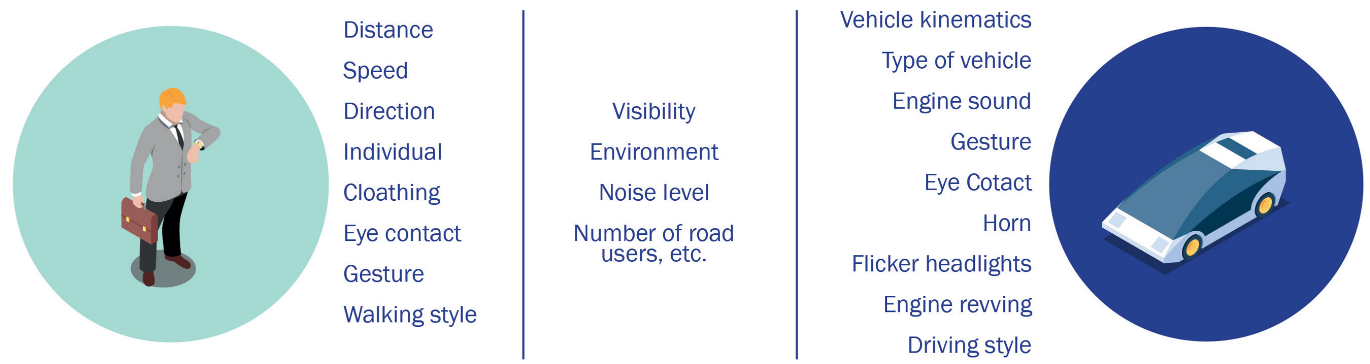


Figure 2.10. A schematic overview of user cuse used by pedestrians when interacting with vehicles. Reprint from "Communication intent of autonomous vehicles to pedestrians", by Habibovic et al. (2018)

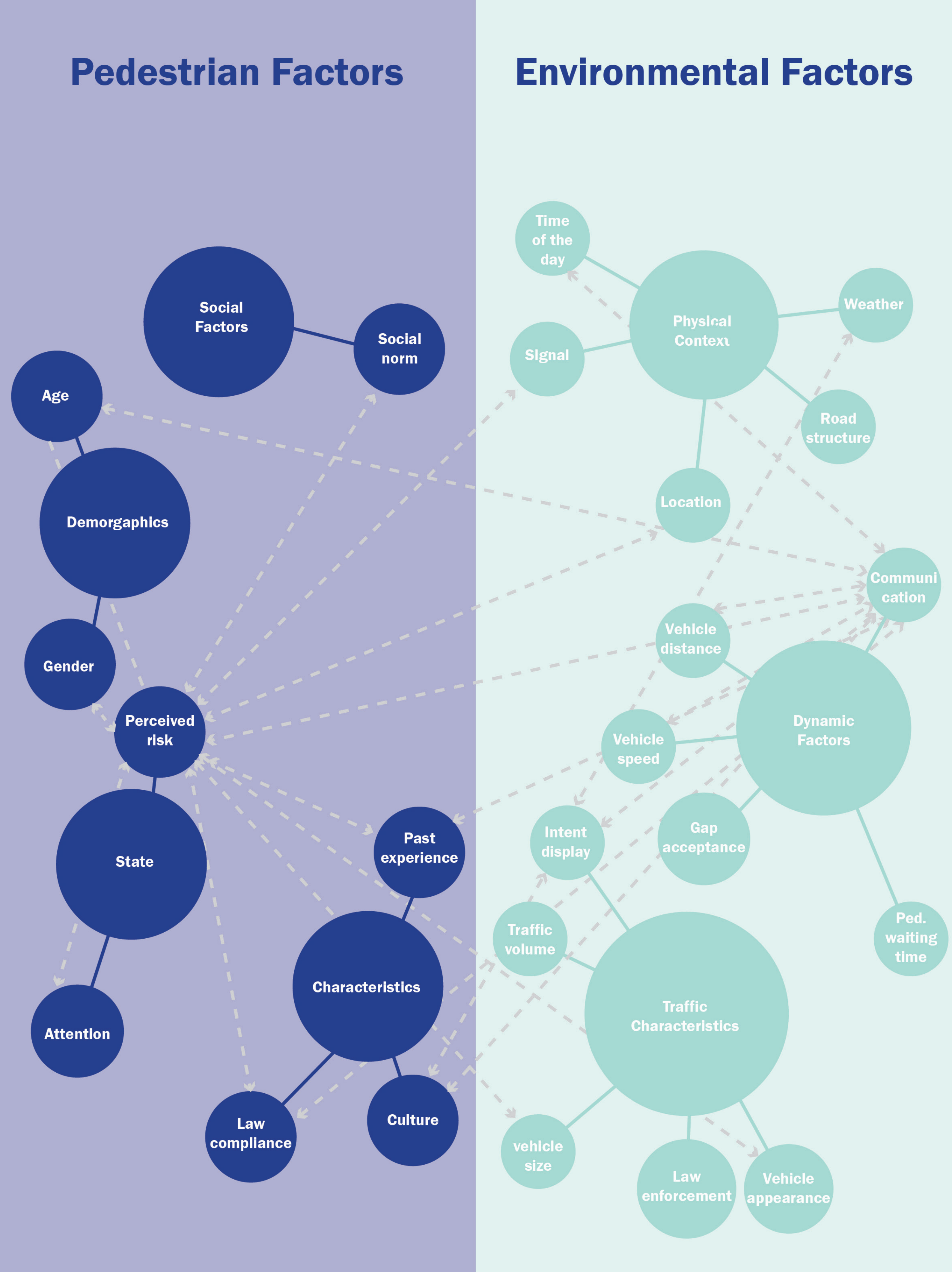


Figure 2.11. A schematic overview of interaction between pedestrian factors and environmental factors in case of autonomous vehicles. Reprint from "Autonomous Vehicles That Interact With Pedestrians: A Survey of Theory and Practice", by Rasuli & Tsotsos, (2019)

Other challenges that VRU face during their interaction with AV are as follows:

Additional Value for VRU: The introduction of AV does promise safety for VRU, but in developed countries with existing road infrastructure (zebra-crossings, overwalks etc.) to provide safe interaction between VRU and AV the additional value creation does not seem like a lot. Raising the question that "Why should VRU change their behaviour to co-exist with AV when the benefits favourable to AV users. (*The insight is from a lecture by Dr. John D. Lee at TU Delft on the topic "Individual and Societal Trust in Automated Vehicles", March 3rd, 2019*).

Social Norms v/s Legal Norms: Social norms do not necessarily need to align with legal norms. This is illustrated by Rasouli & Tsotsos (2019), a speed limit is a legal norm but if the majority of the drivers exceed the speed limit the social norm of driving over the speed limit is created that does not align with the legal norm. These social norms are known to the drivers and VRU who regularly interact with them. However, if an AV is to be introduced a challenge is created because AV will follow strict legal norms and might not be aware of social norms that oppose those legal norms.

VRU Detection: This challenge is more from the perspective of the AV. While conventional pedestrians and cyclists can be easily identified there exist certain cases in which the AV may fail to realise that a VRU is on the road and thus might lead to accidents or an unintended consequence. This scenario will not be prominent in case of highway driving but within suburban localities VRU detection is of paramount importance. This challenge however can be addressed with larger training set data for AV and more naturalistic studies. However, in the meantime solutions need to be developed to (*The insight is from Sasha Arnoud, Director of Engineering, Waymo lecture at MIT Self Driving course, Feb 16, 2018*).

The previously mentioned interaction and challenges VRU face with respect to AV is by no means exhaustive and there exist a lot more literature studying factors influencing this interaction. However, there exists a limitation in the majority of these research, most of the research have studied interaction of single VRU with a single AV. There is a lack of research in the field of interaction between multiple VRU and AV along with AV interaction with VRU in mixed traffic conditions.

Till now we have focused on the interaction of VRU and AV followed by the challenges they face. From the perspective of trust we see that VRU will trust AV when they are able to replicate the interaction that existed with convention vehicles. However, an interesting situation arises due to the behaviour change of VRU especially pedestrians that can potentially harm the introduction of AV within cities. This behaviour change is based in mistrust of VRU towards AV. Autonomous vehicles are designed to keep users safe, pedestrians were seen to misuse this design feature by crossing the road at undesignated locations, because they know an AV will stop for them (Rasouli & Tsotsos, 2019). This brings out a new challenge to the development of AV. Till now we have been focused on the interaction, but this type of change in behaviour raises the question of "How would the behaviour of other roads users evolve with the introduction of AV?".

The above question is an interesting research area, but falls outside the scope of the current project. From the research into the interaction between VRU and AV, we see that communication plays a key role in not only the acceptance of AV but also trusting them. In addition we concur that the influence of the social aspect plays a major role in the acceptance of AV, more than what might be seen in the case of drivers and AV.

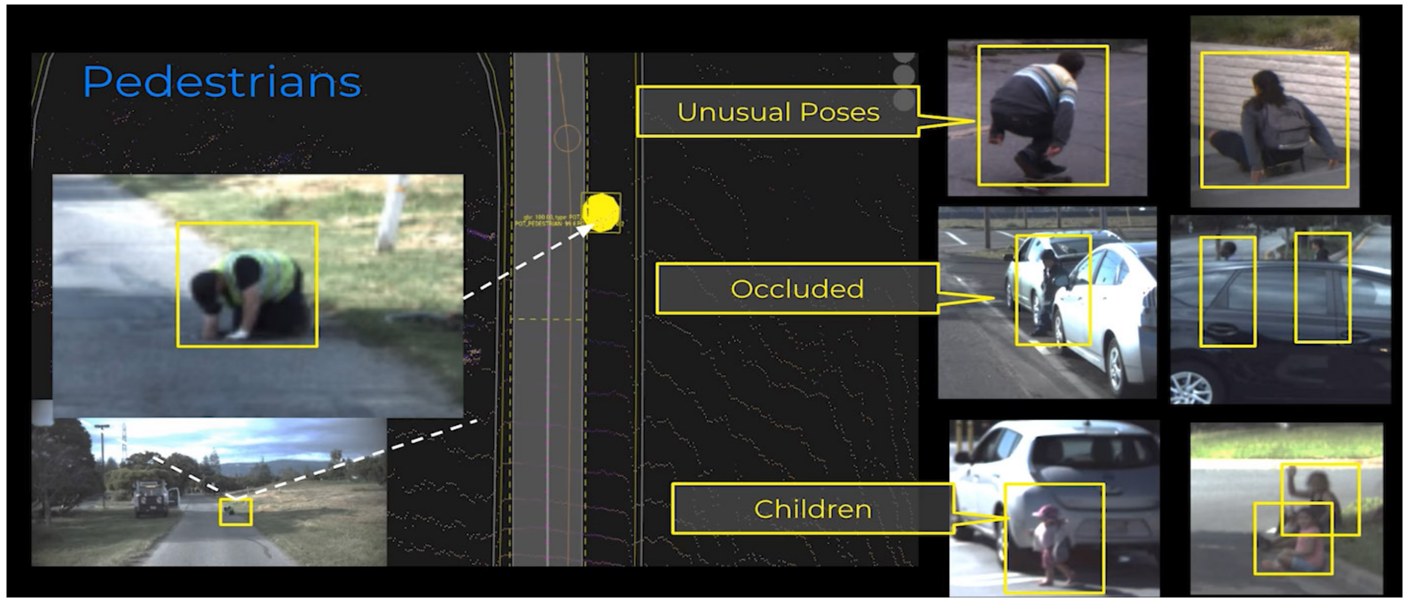


Figure 2.12. Challenge in detecting VRU because of unusual poses, occlusion and other factors, snapshot of presentation by Sasha Arnoud, Director of Engineering Waymo lecture at MIT self driving course, Feb 16, 2018

2.6 | Human Values & AV

Current Research

Human values are one of the overlooked areas when considering designing for autonomous vehicles. While there exists research in the field of designing for values, most of it is directed towards designing for privacy for AV. Values such as universal usability, autonomy, courtesy and calmness that hold relevance in the designing for AV are not discussed.

The focus towards privacy can be attributed to the fact that not only is it a human value, but is also a part of human rights and legislation. Even though there has been research into privacy, most of the challenges it poses to AV are speculative in nature and derived from experience with other technologies. An example of this can be seen in the work of Lim & Taeihagh (2018), who highlighted that breach of private data can lead to dataset biases against people of certain ethnicity or sexuality, harassment through tailored advertisements and marketing strategy and use of data by insurance companies and credit rating agencies. This is not just the case within research, Lee & Kolodge (2019) analysed the response of nearly 15,000 US citizens with respect to their purchase intention for autonomous vehicles, highlighting that common consumers consider an autonomous vehicle as a normal vehicle with a computer attached. This is seen in their responses to the concerns with autonomous vehicles (Figure 2.13), like computers make mistakes, hacking and glitches or error and failures.

While creating a transparent process (Pype et al.,2017) or the use of transparency by design (Langheinrich, 2001) can address the challenges of privacy, we have still seen limited research in the area. A similar case can be seen for cybersecurity as well, Linkov (2019) in his research highlighted that the impact of cybersecurity will be high for SAE level-3 autonomous vehicles and there is a need to research the influence of mistrust on cyber attacks.

We can see that considering human values in the development of AV has the potential to address latent problems and at the same time build trust with the user as seen in the case of privacy by design "*Privacy by design does not promise or aim for complete security or total privacy. What it can achieve is in preventing unwanted accidents and building trust between users and the company*" (Langheinrich, 2001). This is also pointed out by Selbst et al.(2019), stating that, the current approach of "*considering the machine learning model, the input, and the outputs, and abstract away any context that surrounds [the] system*". In simpler terms making a socio-technical challenge into a technical problem results in losing a majority of the social aspects in the process. Thus, there is a need to consider designing for values, value sensitive design and participatory design when considering a socio-technical problem (Aizenberg & Den Hoven, 2020)

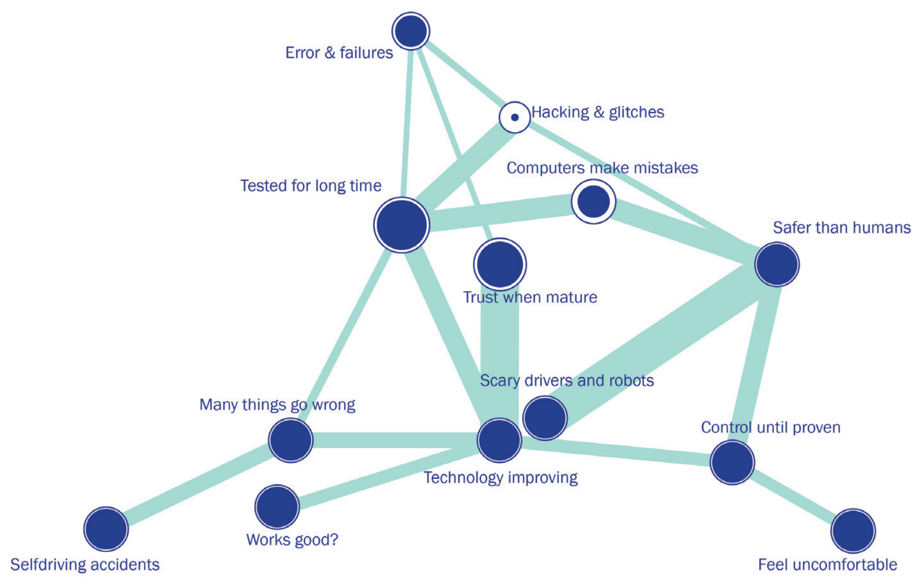


Figure 2.13. Links between topics based on topic correlations. The width of the link reflects the strength of association, and the size of the nodes reflects the prevalence of each topic. Reprint from "Exploring trust Self-Driving Vehicles through text analysis", by Lee & Kolodge(2019)

Changing Values

Designing for human values as discussed above has the potential to address latent challenges. However, there also exists an innate challenge to designing for values, with regard to values changing over time. Vehicles are bought to be used for about 11 years (for Europe). This is a long duration and the values initially considered could have changed or become irrelevant. van de Poel (2018), suggests that values can change due to the subsequent reasons:

- *Emergence of new value*
- *Change in the relevance of the value*
- *Change in how values are conceptualized*
- *Change in how values are specified*
- *Change in the priority or relative importance of values*

To combat the change of values when designing for values, Poel (2018) suggests that new technologies must have the three design features:

Adaptability: The possibility to change the composition or configuration of an artefact or system in order to better perform the original function.

Flexibility: The ability to use the same design in different possibilities.

Robustness: The ability to perform the desired function when the circumstances deviate from the desired conditions.

From the section we can conclude that, while the importance of designing new technologies by considering human values is well recorded in literature. Its implementation in the field of autonomous vehicles is somewhat limited. In addition the challenge of changing values poses a new dilemma for designers and developers of autonomous vehicles. Overall, it is difficult to illustrate how trust as a value interacts with other human values in the context of autonomous vehicles and what role it plays in the acceptance of autonomous vehicles when considering a designing for values approach.

Socio-technical Context

We have come across the term socio-technical on two instances in our discussion so far. The first time it was used to describe the challenges faced in the interaction of VRU and AV, the second time when highlighting the importance of considering the role of human values when designing for new technologies. In this section we explore the meaning of a socio-technical context and the challenges faced in designing for such a system.

A socio-technical system is defined as a "Complex people-containing entity which are conceived and implemented as instruments to serve some purpose" Franssen (2015). The "socio-" part of the phrase does not merely represent the presence of people, but also the fact that instructions, rules and regulations and other social mechanisms are also present to achieve the desired task (Franssen, 2015).

Designing for socio-technical systems differs from designing for conventional all hardware devices in four major ways:

Sociotechnical systems are not designed, assembled or tested from scratch. They are designed and deployed modularly. This is because deploying socio technical solutions are prone to emergent behavior. An example of emergent behavior is the case of misuse of AV by pedestrians discussed in the section of vulnerable road users.

Sociotechnical systems are designed and tested to considerable extent and are conceptualized and monitored from a design perspective. This refers to the fact that the selection of the user of the sociotechnical system is not the responsibility of the designer. The only aspects that are in control of the designer is the technology part of the system.

Users who become part of the sociotechnical system are not considered as hardware components but perform their role as persons/citizens. That means the user is not just another part of the system but is allowed to reflect and analyse this role and the performance of the system.

The role of the user within the sociotechnical system is not just local but global. This refers to the fact that the user is encouraged to reflect on the performance of the system and understand the global consequences of the system.

These differences highlight that a sociotechnical system design is not as simple as that of a fully hardware technology. In addition to this difference there also exists the debate of whether the user within an autonomous system has autonomy to make decisions based on the current condition of the system or if he is just part of the automation (just needs to follow the instructions) (Franssen, 2015). While there exists more literature on designing for sociotechnical systems. The current differences as discussed above provide a basic understanding of the term and how it differs from conventional design of full hardware systems.

2.7 | Product Development Process

Introduction

The scope of the project is focused towards the autonomous technology development team, for this reason it is imperative to explore the process of design and the interaction between the stakeholders that were introduced in the initial section of the report. We begin the section with an introduction towards the various stakeholders present in the product development process and then move towards the product development process and the stages present. We conclude this section by drawing out insights from the product development process that will influence our design toolkit.

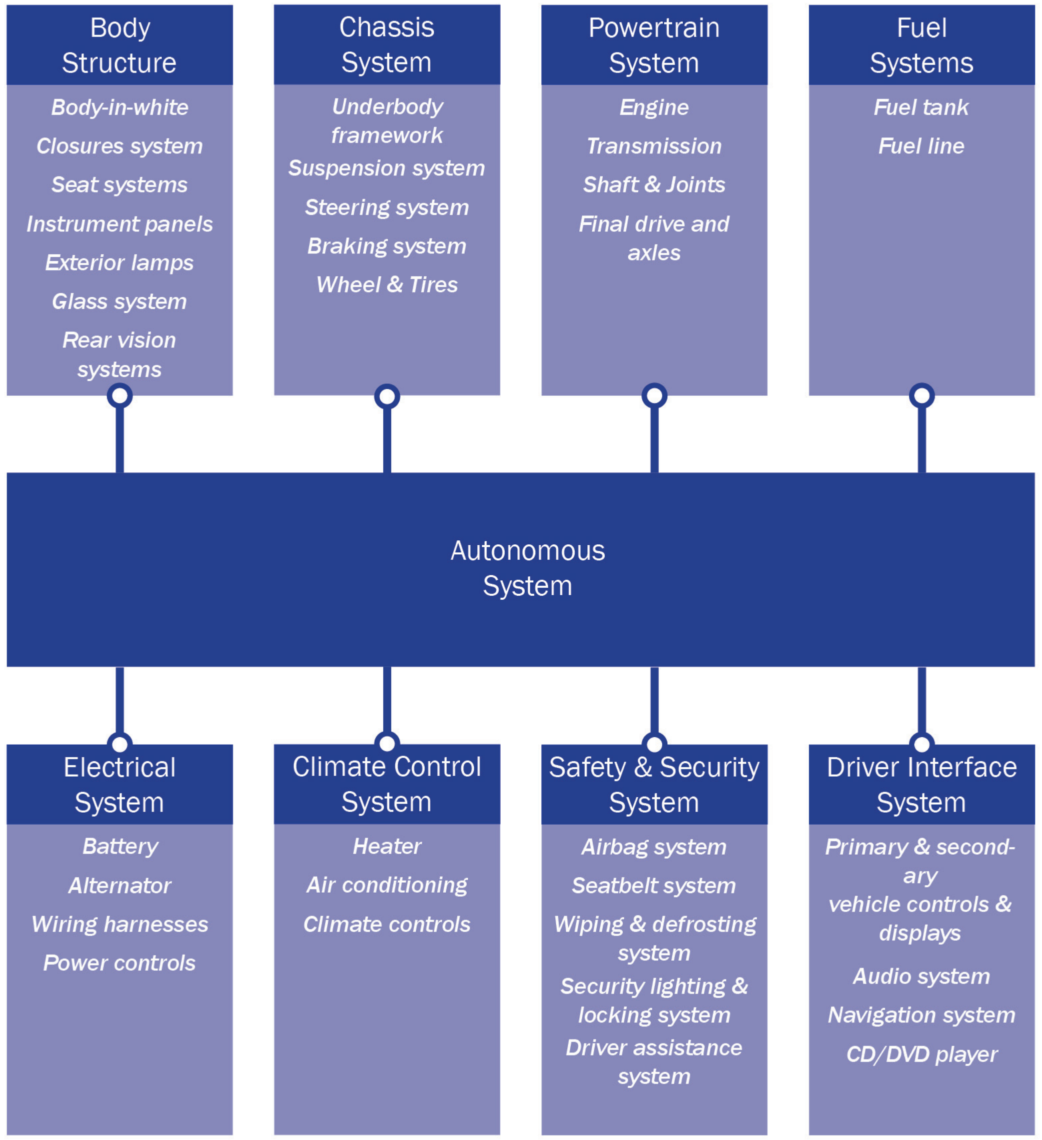


Figure 2.14. The various systems and subsystems present within an automobile as presented by Bhise (2017)

Stakeholders

As described in the introduction we classified the stakeholders into three categories, the primary stakeholders, secondary stakeholders and tertiary stakeholders. The primary stakeholders is the autonomous technology development team. The secondary stakeholders are the other systems that are present with a vehicle. These systems have been placed within the secondary stakeholder tier because they influence the primary stakeholder during the product development process. The tertiary stakeholders are departments/functions/systems that come later into the product development process, the manufacturing process, sales, distribution and finally the end consumer.

Primary Stakeholder

The autonomous technology development team consists of the subsequent sub-systems: Hardware engineering, Software Engineering,System design and Industrial Design. This team composition is a broad representation of the subsystems that are present within the autonomous technology development team. However, the number of individual employees present within each of these subsystems is not fixed and can vary from organisation to organisation.

The Autonomous technology development team is directly connected to each of the secondary stakeholders (other systems). This is because an autonomous system exerts control over other systems of a vehicle when engaged. Example: when adaptive cruise control is engaged the autonomous system takes over the steering control (part of the chassis system), the brakes (braking system), the acceleration of the vehicle (powertrain system) and HMI (Driver interface system). Thus, the secondary stakeholders play an important role in realizing the capabilities of an autonomous system and proper communication needs to be maintained between the stakeholders during the product development process.

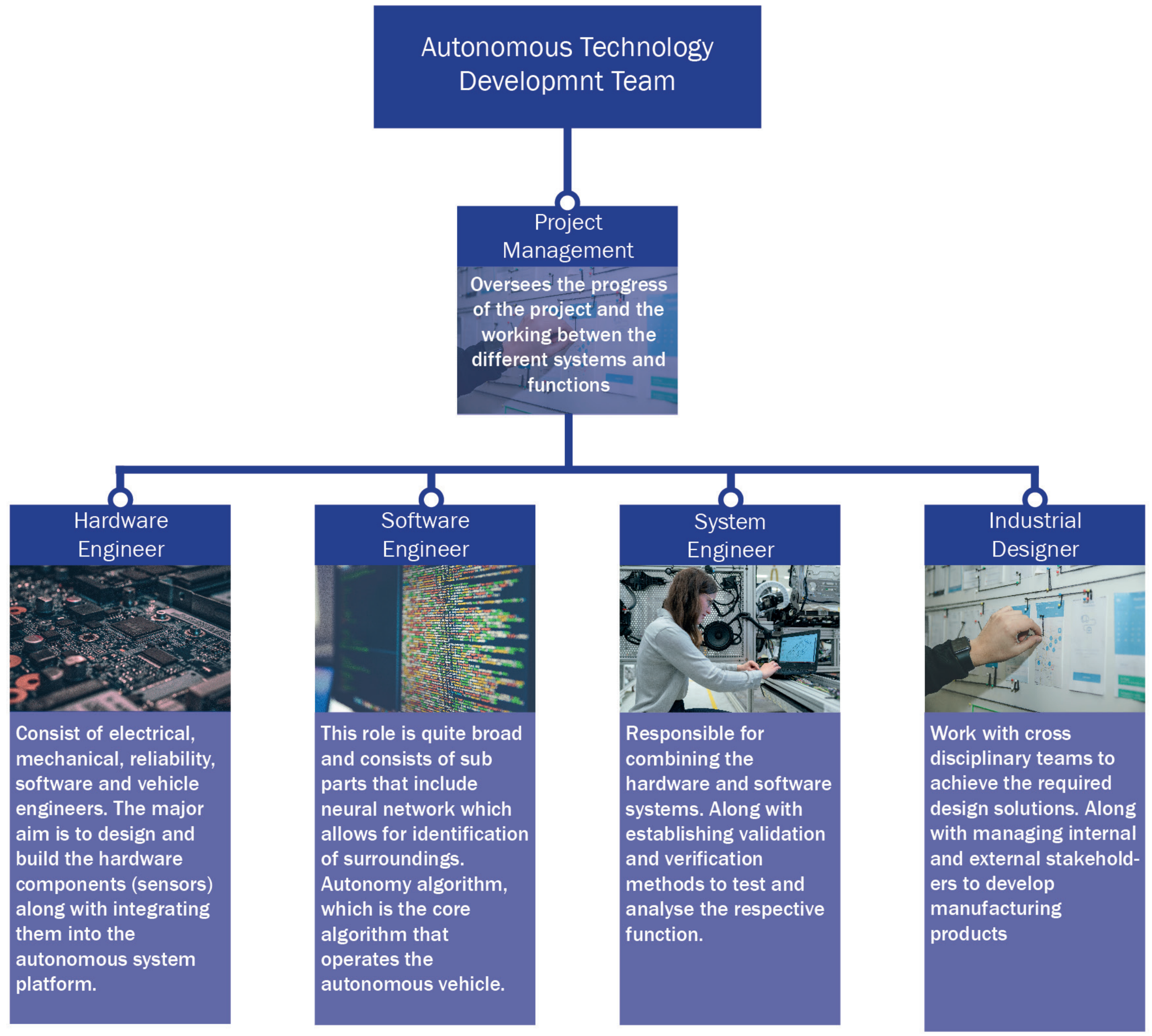


Figure 2.15. Breakdown of the autonomous technology development team. The team structure is based on the team composition as represented within Waymo, Tesla and Mercedes Benz.

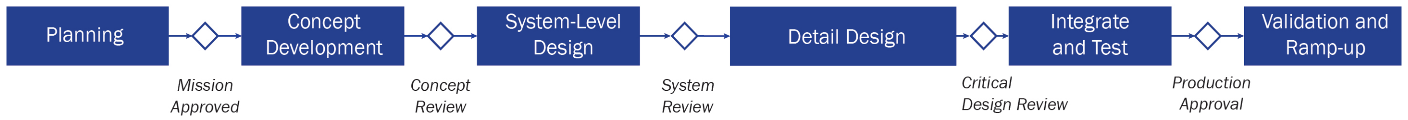
Automobile Product Development Process

There does not exist any standardised automobile development process, while the major stages and process of the development process remain the same, organisations have bespoke product development processes that suit their needs and demands. For the current project we will consider the product development process proposed by Ulrich *et al.*, (2019) in the book “Product Development and Design” and supplemented by Bhise (2017) book on automobile development titled “Automobile Product Development: A System Engineering Implementation” and Sorensen (2016) book “The Automotive Development Process: A Real Options Analysis”.

The conventional product development process has six phases: Planning, Concept Development, System-Level Design, Detail Design, Integrate & Test, Validation & Ramp up. The product development process for an automobile is considered a complex process (Ulrich et al., 2019) and thus has some differences. The main difference is in the Detail Design Phase, as there are a number of different systems within an automobile all of them are designed simultaneously in parallel. This means that there are a lot of interdependencies between the all the systems development process.

The main takeaways from the product development process that will need to be considered during the design phase of the project is the interdependency of the systems and the importance of communication between the various stakeholders. In addition, the autonomous system along with the other systems must meet the vision and mission statement on which the vehicle is designed.

Conventional Product Development Process



Automobile Product Development Process

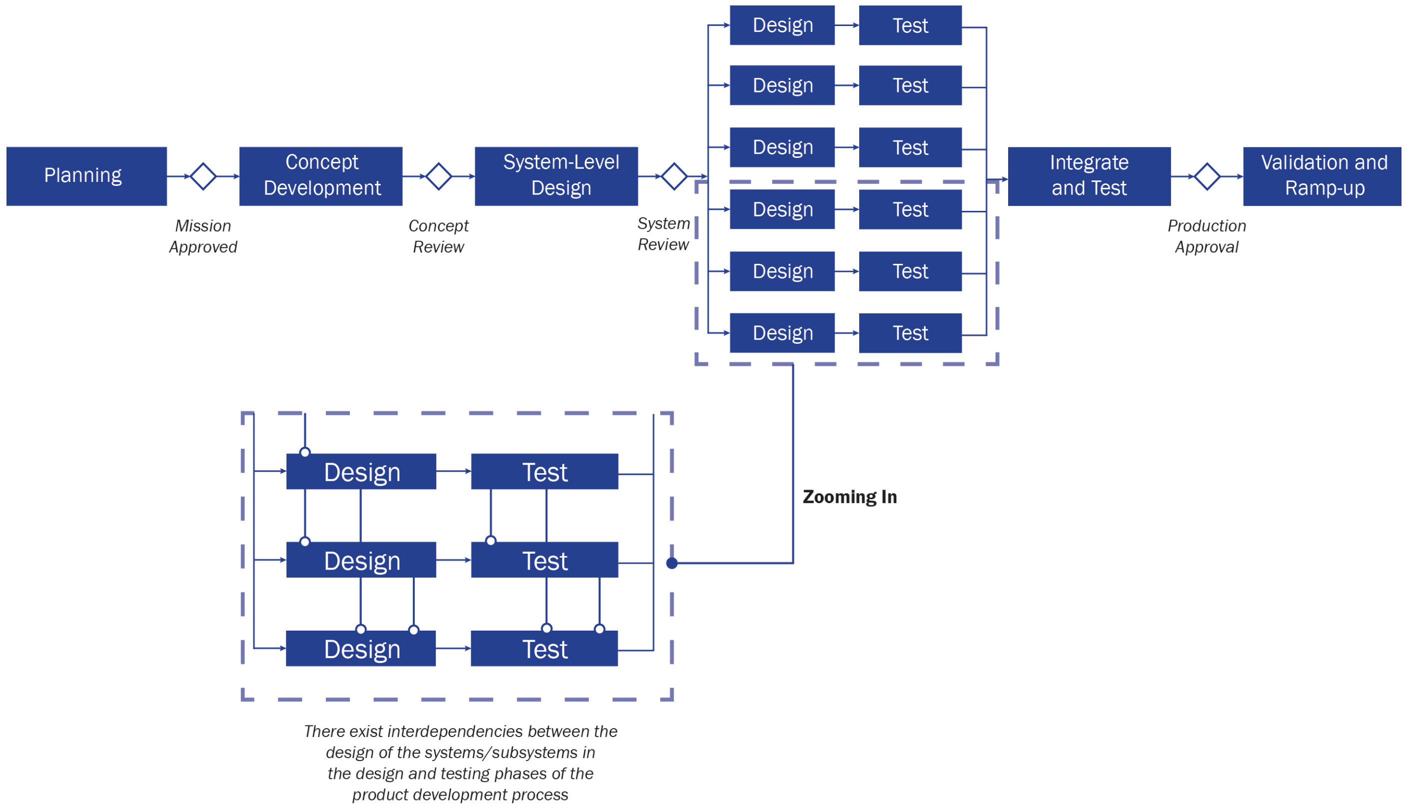


Figure 2.16. Product development process for automobiles, based on the process flow diagram presented by Ulrich et al.(2019)

Key takeaway from the Product Development Process

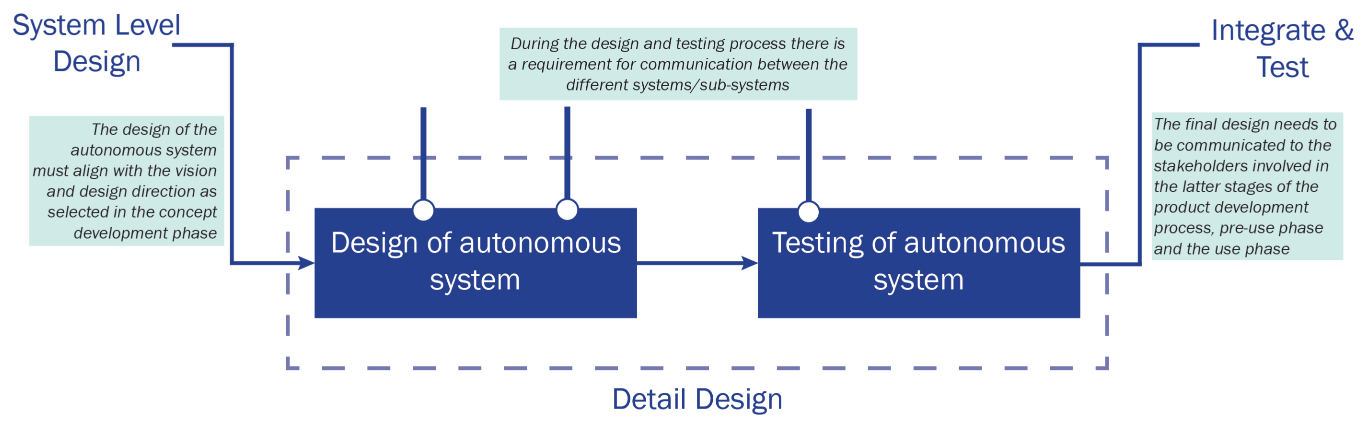


Figure 2.17. Key takeaway highlighted in light green from the product development process, specifically in the details design phase of the product development process

2.8 | Conclusion of Literature Review

Brief Recap

The literature review started with the understanding of trust as we understood the process of trust formation and the factors that influence trust as proposed by Hoff & Bashir (2015). Subsequently we were introduced to the concept of calibrated trust and the current research in the field of calibrated trust. We then shifted towards understanding the role of drivers and the influence of autonomous vehicles on the driving process. We were introduced to the ironies of automation and the current challenges faced by users. We then moved our focus towards another type of user in the form of vulnerable road users. Exploring the interaction between VRU and AV along with the challenges highlighted in literature. Humans values was the next section that was explored along with socio-technical systems. We concluded the literature review with an understanding of the product development process and the composition of the autonomous technology development team.

Key Conclusions

- *Trust can be divided into three parts: Dispositional trust, Situational trust and Learnt trust. However, design features in automation only influences the dynamic learnt trust part.*
- *Trust in automation is not constant and is continuously changing. In addition one cannot measure trust using a single scale value*
- *Calibration of trust is in essence the calibration of the users mental model, which is based on understandability, predictability and reliability of the autonomous system.*
- *Current design and research into calibrated trust is focused towards HMI systems which again focuses on the dynamic learnt trust. However, there considering situational trust and initial trust have the potential to aid in calibration of trust*
- *Autonomous systems were introduced to increase the capabilities of the driver but also seem to increase the taks demand of the driver. Thus, a new balance needs to be created between the combined capabilities of the driver and autonomous system against the task demands.*
- *The biggest challenge to AV is the introduction of autonomous systems that are not fully robust and reliable but are complex for users to intervene in a meaningful manner.*
- *Trust between vulnerable road users and autonomous vehicles is based on implicit communication, in a social setting making the design of autonomous vehicles a socio-technical problem instead of a rigid dynamic problem.*
- *When designing for trust in AV, there is a need to consider that human values can change through the life cycle of the AV.*
- *Designing technology in a socio-technical system is different from designing a purely technological system and the differences must be incorporated within the design process*
- *The product development process for automobiles is a complex process with a number of different systems. Each system is designed separately but there exist interdependencies between the systems that must be catered for in the design process.*

"If I have seen further it is by standing on the shoulders of Giants"

Issac Newton

Photo by Nice M Nshuti on Unsplash

Part 1 I Chapter 3

Qualitative Research

The chapter covers the qualitative research that was conducted during the project. 16 interviews were conducted within the scope of qualitative research including experts from the field of autonomous vehicles, trust in automation, vulnerable road users and responsible innovation. Thematic analysis was employed to analyse the interviews. Prominent results of include the trust formation process of a user, barriers to calibration of trust and the changing role of designers.

3.1 I Research Setup

Introduction

The empirical study is focused towards creating a coherent overview of the autonomous vehicle landscape and to get a better understanding of the role trust plays in the current design process. To achieve these two goals semi-structured interview guide methods were selected. In addition to the ability to conduct interviews during quarantine, semi-structured interviews allowed for interviewing a diverse range of experts while at the same time providing a systematic and comprehensive method to gather data (Patton, 2002).

Aim of the Interview

- *Insight into current research and methods used in practice with regard to calibration of trust*
- *Prominent challenges faced in the acceptance of autonomous vehicles and the role of trust in these challenges*
- *Exploring the design process of engineers and developers The role of human values in the design process*
- *Challenges faced by OEM of autonomous vehicles*
- *Role of designers in the development of autonomous vehicles*

Planning the Interviews

The major themes discussed in the literature research were chosen as the starting point for planning the empirical study. The themes were used to define four topics to conduct empirical study: *Autonomous vehicles, Trust in Automation, Vulnerable road users* and *Responsible Innovation/Human values and AI*. Once the four major themes were selected the interview guide was created based on the aim of the interview and the associated topic, based on the guidelines as proposed by Patton (2002). In total four separate interview guides were created that represented the four major themes that were identified in the literature review.

Selecting the participants

The sampling strategy chosen for the project was strategic and purposive (Miles et al., 2013), in association with the identified four research themes. The selection of the participants was done in three steps. The first was creating a list of researchers from the research papers published in the specific domain. This was followed by searching for participants within the multiple faculties at TU Delft and TU Eindhoven. Due to the COVID-19 lockdown, certain experts were unable to confirm their participation for this reasons, a third step was employed to access experts. Sampling through snowballing was employed (Patton, 2002). In total 16 participants were confirmed, 4 for each of the four different topics that had been defined. It should be noted that not all participants were not strictly associated with one single topic but had expertise in multiple topics. In such cases the use of the semi-structured interview allowed for exploring additional topics with these participants because of their varied expertise.

Conducting the interviews

All interviews were conducted by Skype or Zoom because of the pandemic. The interviews were audio recorded. The duration of the interviews was between 35-55 minutes. Following the interviews the audio was transcribed using Otter.ai (a website for transcribing audio and video files). A second round of manual transcription was performed to check for any mistakes or errors in the transcription.

Figure 3.1. The four areas of expertise that were interviewed for the qualitative research during the project

3.2 | Data Analysis

Overview

After the completion of the transcription, the interviews were analysed by using Atlas.ti Version 8. The four topics were analysed separately and then interrelations were identified within the four separate topics. Thematic Analysis was chosen to analyse the interviews, this is because it allows for identifying topics that integrate into higher-order key themes (Spencer et al., 2014). The first step in the analysis was familiarizing oneself with the data, this meant reading the transcripts once and simultaneously taking notes of broad themes that were discussed during the interview. This was followed by the first round of coding, *in-vivo* coding was performed during this round. The second round of coding performed was focus coding.

Having completed the focus coding, themes were identified within the codes. There was no specific method of selecting themes, they were selected based on the significance to the research question. Having completed the creation of themes, these were reviewed once to check if the initial codes fit the chosen themes (Maguire & Delahunt, 2017). The last step in the data analysis was to triangulate the data, for this theory triangulation was performed to triangulate the themes identified with insights gathered from literature.

Figure 3.2 Process for performing thematic analysis of qualitative data as presented by Braune and Clarke (2012)

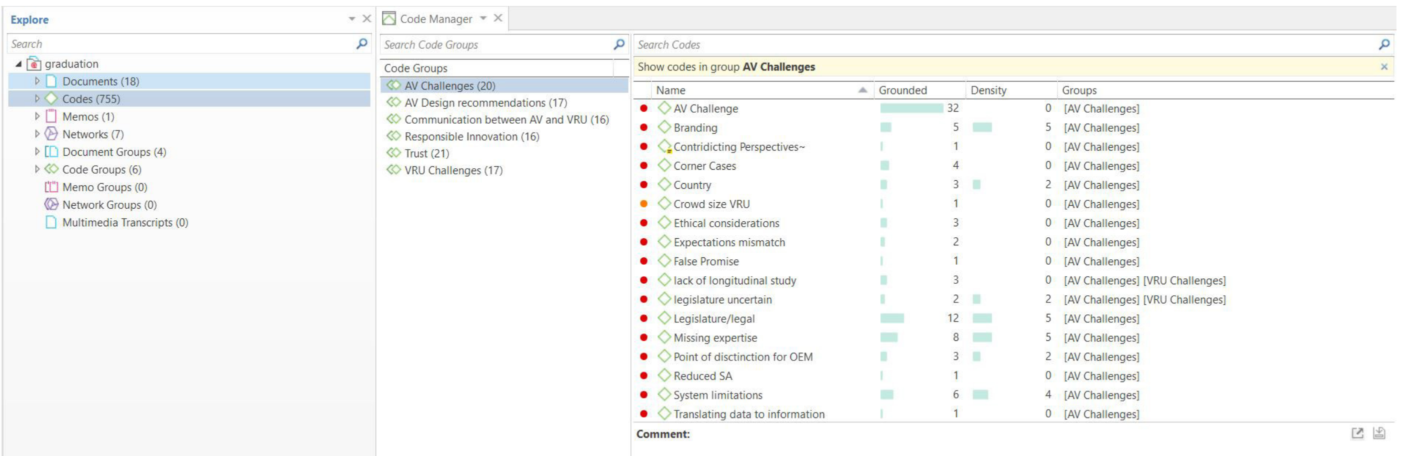


Figure 3.3 A snapshot of the codes and focus codes created during the thematic analysis of the qualitative data on Atlas.ti Version 8

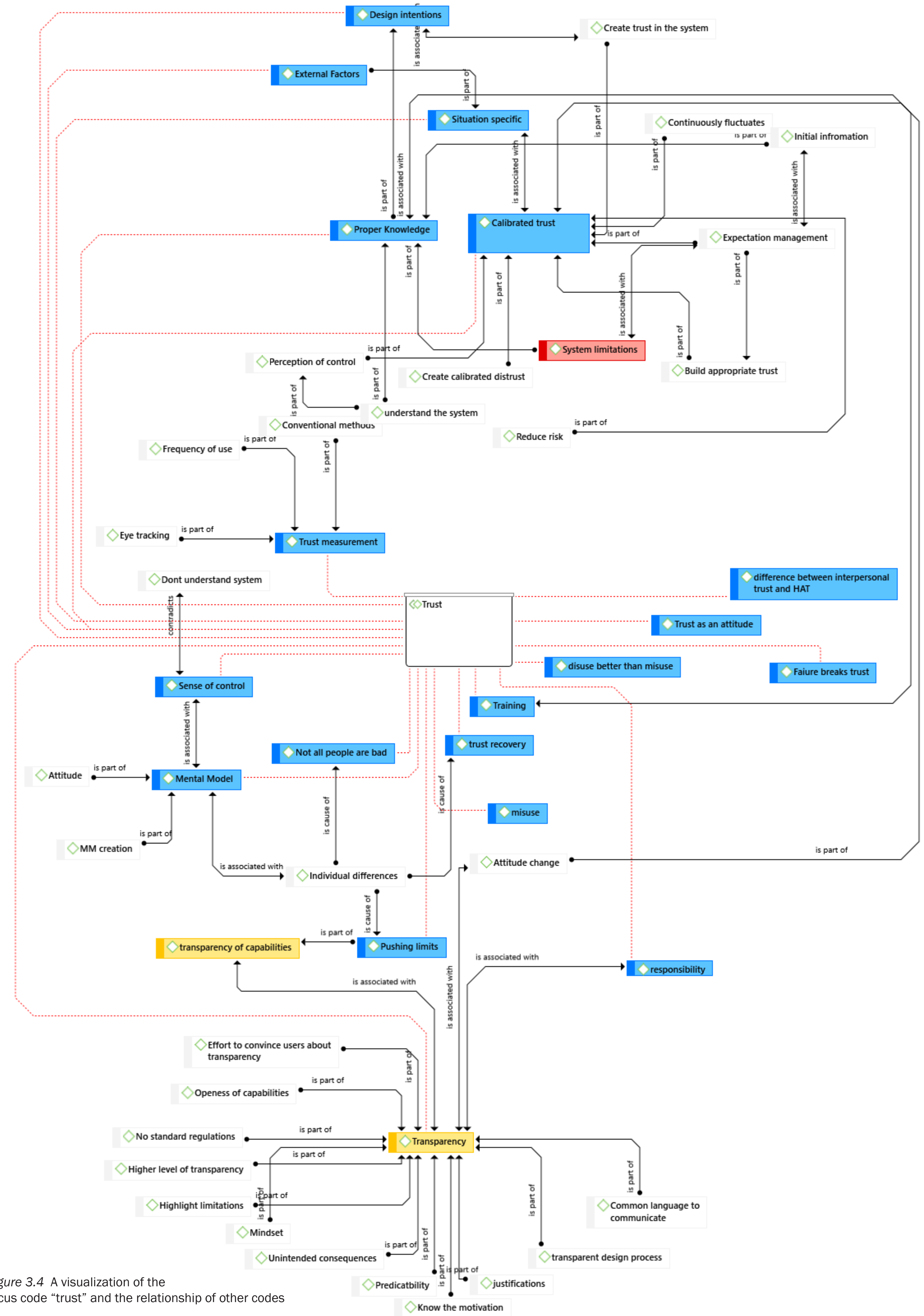


Figure 3.4 A visualization of the focus code "trust" and the relationship of other codes

3.3 | Result

The results of the analysis have been categorised into three parts. The first is the discussion of how trust is developed when interacting with AV. The second is barriers that were highlighted in the process of developing autonomous vehicles and the last part is a collection of prominent points highlighted during the interview.

Trust Development in AV

Figure 3.5 shows how calibrated trust is developed assuming novel interaction of users with autonomous vehicles. The horizontal axis is the timeline starting from the product development process followed by the pre-use stage and finally the use stage. The product development stage starts from product planning and ends at production and ramp up. The pre-use phase begins after the production stage and ends at the point the vehicle is sold to the user. The use phase consists of the complete duration of use of the vehicle until discarded. The vertical axis divides the trust between the company and the autonomous vehicle. The trust level is a semantic representation, ranging from mistrust to distrust. At the center there is the calibrated trust region. The grey zone represents the zone within which trust can exist (since trust cannot be measured we cannot pin-point a specific value for it, thus the light green zone represents the possibility of trust existing in that region). We see that in the product development stage the trust can vary across distrust and mistrust, this is based on the previous knowledge the user has with the company along with his disposition trust aspects such as age, gender etc. As we move into the Pre-use phase we see that the region of trust begins to converge towards the calibrated trust region. This is because as the user gains more information about the autonomous vehicle their expectations start to align with the capabilities of the system. As the user enter the use phase we ideally would like the trust to match with the calibrated trust levels.

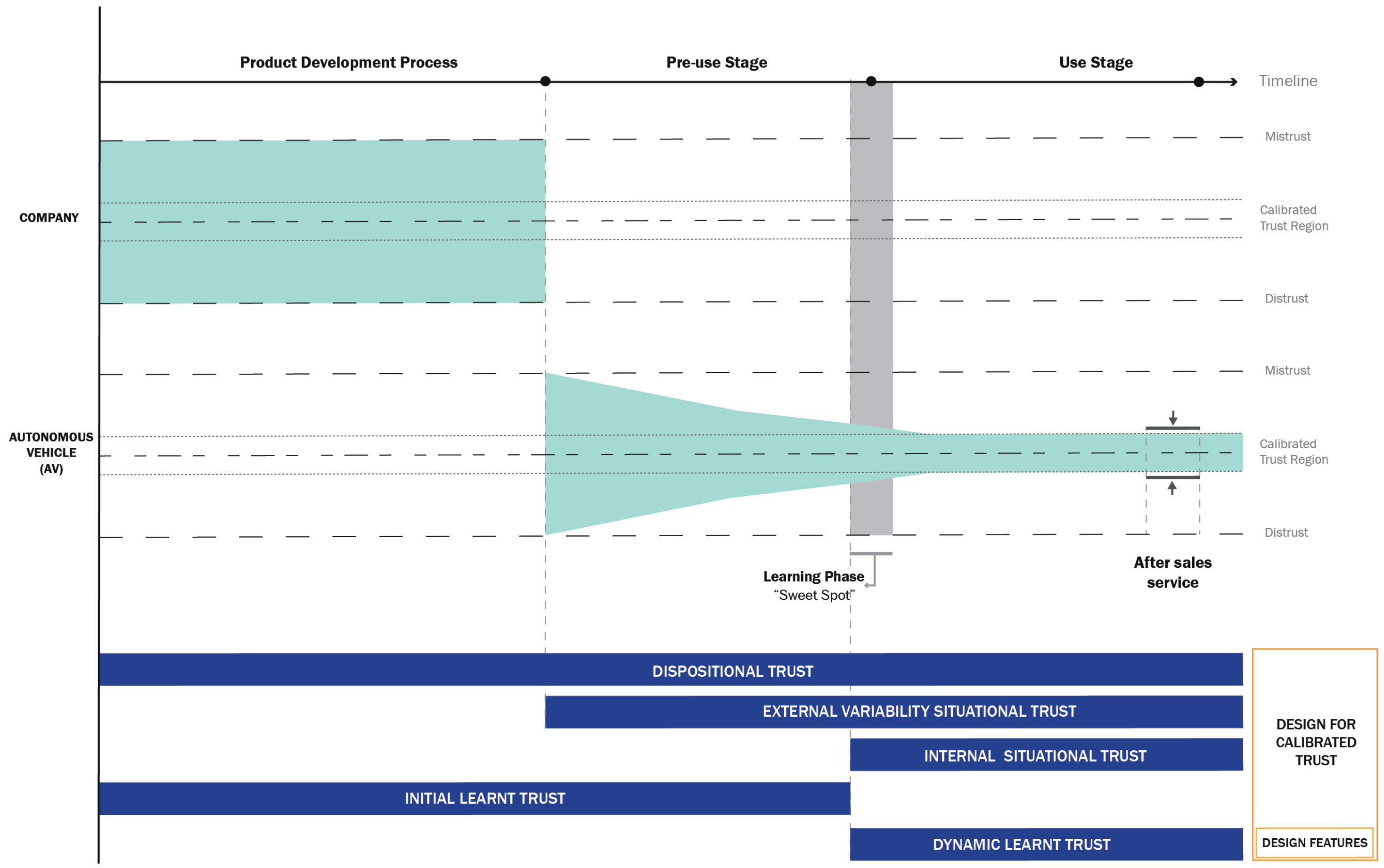


Figure 3.5. The ideal trust development based on novel interaction of user with autonomous vehicle in the case of purchasing an Autonomous vehicle.The light green region within the image represents the possible trust location at that given point of instance.

The grey strip represents the sweet spot to ideally reach calibration of trust with regard to understandability and predictability. It is important to point out that calibration of trust is not a constant state but is constantly changing based on the context. The reason for highlighting a sweet spot is because we want to achieve calibration of trust with regard to reliability in the use phase and not have to calibrate the understandability and predictability. An optimistic sweet spot for calibration would be within the pre-use phase i.e. the users understandability and predictability are calibrated before the purchase of the vehicle. A pessimistic sweet spot would be in the use phase, where the users understandability, predictability and reliability are all calibrated in the use phase.

The reason for identifying the use phase as a pessimistic calibration spot is because the responsibility of calibrating the understandability, predictability and reliability fall to the HMI system. This is an unfavourable situation in comparison to calibration in the pre-use phase where there are multiple information sources for calibration. An analogy to understand this can be the concept of integration and differentiation. The pre-use phase is like integration where the more information and experience we have the more accurately we can calibrate trust. The use phase is more of a differentiation function, in the sense that trust is continuously changing based on the context of use. This makes it a bigger challenge to achieve the right level of calibrated trust through the use of HMI systems only during the use phase. Due to this we see that by achieving a certain level of trust calibration within the sweet spot will allow for a more coherent integration of HMI systems into the designing for calibrated trust.

If we were to compare the process of calibrating trust for novel use purchasing an AV with other forms of mobility such as AV in shared mobility (Uber) or as a part of mobility as a service (MaaS). There exist certain differences. In the case of shared mobility, trust is a two way process with the user and the driver both placing trust on each other (Etzioni, 2019). In the case of a fully autonomous ride sharing vehicle the trust would again be both ways between the user (the vehicle is safe, the AV will choose the shortest travel time) and the company owning the vehicle (user does not damage the vehicle). In addition to this trust of users towards a shared mobility might be a representation of the distrust towards other forms of mobility (Etzioni, 2019). Thus, we see that the trust with relation to the capabilities of the autonomous vehicle is but a part of a larger picture. A similar difference can also be seen with MaaS.

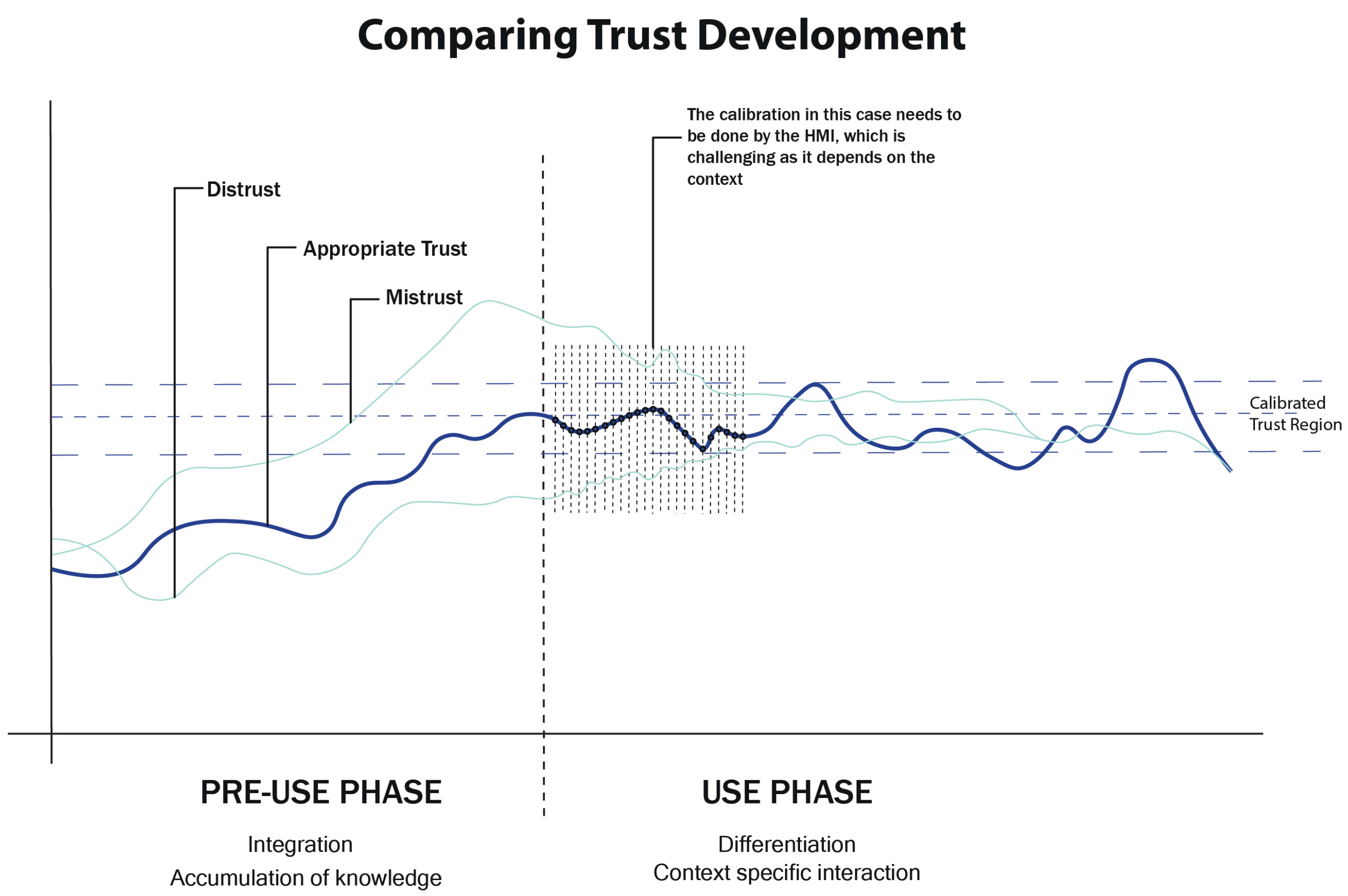


Figure 3.6. Comparing the development of mistrust, distrust can calibrated trust between the pre use phase and use phase

Barriers in the development process of AV

A number of challenges/barriers were highlighted through the various interviews that companies are facing or might face in the development of autonomous vehicles. While some of the challenges have been discussed previously in the literature, this section highlights prominent barriers.

Common Language for communication between stakeholders

Communication is one of the most fundamental barriers to calibrated trust, if an automation cannot communicate its capabilities to the end user and the end user fails to communicate his expectations to the design and development team, we will not be able to reach calibrated trust. The analysis of the interviews shows that there exist three parts to this communication. The first is communication within the company/organization as highlighted by these quotes, "*data scientists have this huge lack of communication, they cannot communicate their epistemological to people that do not understand their concepts. So communication is really important, also managing this problem*" and "*So you need need to have some kind of coherent view, you need to have the same intention from a whole company in order to design a system that you can interpret in a very, in a very coherent way*".

The second part of communication is communication between the company and the end users. This traditionally is achieved by advertising, visiting dealerships, online platforms, etc. The importance of this is to highlight the capabilities of the system "...The purpose of the system and the other part is basically *that the the people behind the automation also are very clear what they mean with their system what how the system should be used and what is safe usage of the system...*" along with the limitations "*So I think they're the understanding of that the user should be aware that limitations are that they may be limitations in certain circumstances so that they can be more aware at least and monitoring the system*" and lastly the responsibilities the user has when using the autonomous vehicle "*...if so so so like a discrepancy in your understanding of the system will lead to like you not being able to take your responsibility at some moment..*". This is also corroborated by the work of Yang, Han & Park (2017), suggesting a more human centered approach to defining the SAE levels of automation to reduce communication gaps.

The last part of communication is between the end users and autonomous vehicles. For the user within an AV this corresponds to the design of the HMI system and for other road users (RU) and vulnerable road users (VRU), it includes eHMI and maneuverability of the vehicle.

The three parts of communication can be seen in succession, from communication within the organization during the development of the AV, communication with end users occurs during the pre use phase of the AV and lastly the communication between AV and users in the use phase. There is a need to create a coherent system of communication to not look at these three parts individually but as a collective whole.

> *"I mentioned the establishing a language in which you can communicate with people across different disciplines. And that's not an easy thing to do. And that takes time. But it will be really essential to do that. So, I think that's, that's certainly a big challenge that needs to be overcome with time"*
>
> -Expert in Designing for Human Values

Solutionism Approach

Solutionism refers to the belief that technology has the capability to solve any challenge. It is the polar opposite of a socio-technical mindset in which we accept that not all challenges within a given context can be solved by the intervention of technology. A solutionism approach to problem solving was cited as a common problem when working with data scientists and researchers. This approach was believed to act like blinkers placed on horses, narrowing the field of view of the problem.

Evolution of user behaviour

The introduction of autonomous systems into vehicles is believed to create a change in the user behavior and in turn user needs. One of the most prominent changes was highlighted by Bainbridge (1983), in which he stated that the new role of user will be to monitor the automation for anomalies and alarms, whereas research shows that humans lose attention and psychological awareness when asked to monitor. This was also highlighted in the interviews, "*Less sensitive in the sense that if the car does, like, go slightly to the left or slightly to the right, it does not affect that much because you don't really see it or feel it*".It was also pointed out that that users speed up when using automation, "*So what happens is that that's a study found out is that people actually speed up on their own we provide that type of assistance*". Other prominent points discussed were skill degradation of the user, unaware of automation mode vehicle is engaged in the vehicle and from a manufacturers point of view the motivation of the user to drive an autonomous vehicle. The points discussed also correspond to the lumberjack effect (Endsley, 2019) "*The higher the level of automation becomes, the more people will trust the automation and are likely to fail in taking over manual control during critical situations*" and Ironies of Automation as discussed by Baindridge (1983). This change in behaviour is not only seen in the interaction of the driver/user with AV but also with VRU and AV, leading to misuse of AV in certain cases.

Key Themes

Branding and Point of Distinction

Branding play an important role in the way users perceive autonomous vehicles. This also includes naming of the autonomous technology as highlighted by the quote "*And then some, some nice words that look nice, like I don't know what there are, like, I think 20 different terms for adaptive cruise control*". This quote highlights two overlooked barriers for users, the first is incorrect representation of technology by their name (Senator Edward J. Markey recommended name change from autopilot to a more appropriate term after the Tesla crash in 2018) and the second is multiple names and definitions of the same tvechnological features by different car manufacturers to match their brand values, in this case adaptive cruise control.

The barriers at first sight might seem to originate from branding but it is rooted in the development of autonomous technology as stated by an interviewee "*So the challenge is that it's like every company will be reinventing the wheel again and again and again, because I don't think Tesla is going to share their information*". Each company is trying to carve out a portion of the autonomous vehicle market at the same time maintaining their brand values. The following quote highlights this "*So they put a lot of effort on showing that they stand out in that in that way in that particular way. So they don't want that they're worried that if automated vehicles are all supposed to behave within certain limitations and restrictions that every BMW also starts to drive like a grandma's car, and then the whole their selling point is there. USP kind of gone*".

We can conclude from these statements that there needs to be a connection between the development of autonomous technologies in conjecture with the existing brand values without compromising the requirements of the end user. Similar outcomes were seen in the research by Lee & Kolodge (2019), specifying that the relationship between the user and the brand play an important role in the formation of trust.

> *"So they put a lot of effort on showing that they stand out in that in that way in that particular way. So they don't want that they're worried that if automated vehicles are all supposed to behave within certain limitations and restrictions that every BMW also starts to drive like a grandma's car, and then the whole their selling point is there. USP kind of gone"*
>
> -Expert in Vulnerable Road Users

Transparency

Transparency is seen as one of the methods for addressing certain barriers (understandability, predictability and recognition of biases) that pertain to AI. Interestingly, none of the participants believe that there currently exists a truly transparent system. The reason for this is explained as to achieve a transparent system there is a need of a transparent design process "*...So one is that the design process itself needs to To be transparent, in the sense that you can see how design choices were motivated...*" and the ability to justify the decisions made by the the AI system "*The second aspect of transparency that I often consider that has to do with you know, what you often see in the context of AI is referred to refer to explain ability or justification of automatically made decisions*".

The above two are considered as the main drivers for developing a transparent system, but there exist additional factors that were also pointed out during the interviews, including a change in mindset towards designing with AI, understanding unintended consequences and development of standards/regulations that have an influence on the transparency of the autonomous system.

Role of a Designer

During the course of the interviews, a number of design recommendations were suggested along with the role designers can play. This section discusses some of the prominent design recommendations that were discussed.

Design AV like a system: Instead of looking at autonomous systems as an additional part to an existing vehicle, we should consider looking at how autonomous systems influence other systems and design it as a collaborative system, "*you (designer) also need to go out when the when the, when the car has been completed, the design of the whole car is completed, you need to go out also to see how the whole car is interpreted together with your system. So you get the whole, like, what does the car actually say to you*".

Improvement not perfection: This can be seen as a warning to not fall into the solutionism trap and try to create a perfect system. We need to approach the design in an iterative approach and consider an AV as a living lab that allows for continuous improvement. This insight ties into not only the socio-technical approach to design but also addressing the theme of evolving user behaviour and building a transparent system.

Critical to Design: It is important to be critical to the design decisions, this ties into creating a transparent system as it allows to understand the assumptions and choices on which design decisions were made.

User Needs and values: A lot of importance was put on the human factors aspect of design, it was also mentioned as one of the prominent areas within autonomous vehicles that seems to be lagging behind along with legislation. The interviewees believed that not only do we require more insight into designing for human factors but it is also important to verify the assumptions made to corroborate with the needs of the end user.

The role of designers also seems to have changed with regard to AV. Designers are expected to provide a more holistic overview of the process. In addition, have a proactive approach towards addressing the needs/values of the stakeholders. Bridging the gap between subjectivity and objectivity challenge when working with AI systems and lastly aiding in creating a common language to communicate across all stakeholders.

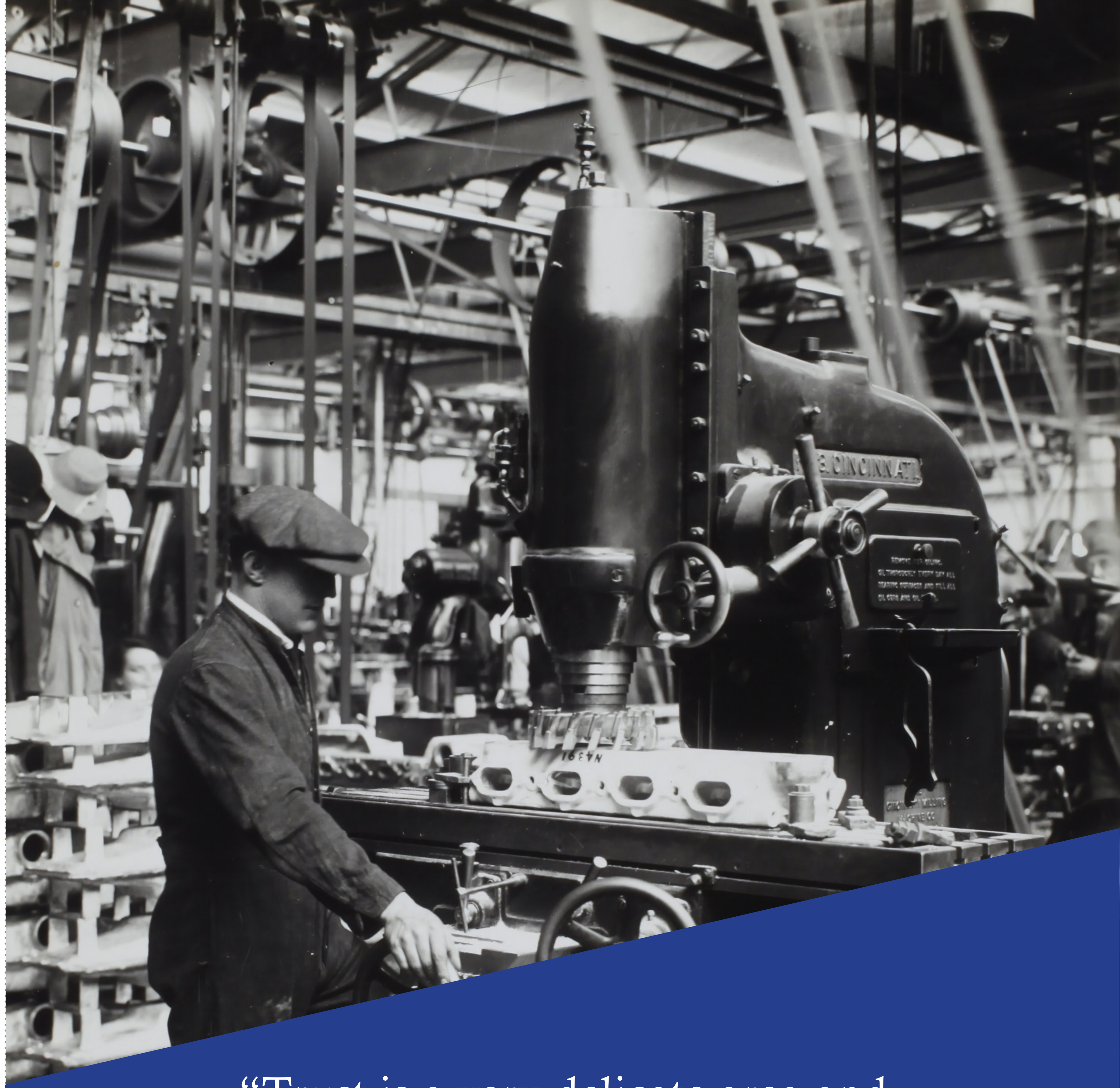
> *"you (designer) also need to go out when the when the, when the car has been completed, the design of the whole car is completed, you need to go out also to see how the whole car is interpreted together with your system. So you get the whole, like, what does the car actually say to you"*
>
> -Expert in Design for trust in AV

3.3 | Conclusion

The interviews conducted as part of the qualitative research process were successful in answering the initial question framed. Certain insights were in line with the literature research such as the importance of a socio-technical approach to design, the role of communication in calibration of trust. Certain other insights were unexpected and interesting to explore especially the role of a designer in designing for calibrated trust. While the analysis of the interviews provided a more concrete understanding of the challenges in designing for calibrated trust, there still exist multiple avenues in which the project can be taken forward. As we move towards the reframing of the design question in the subsequent section it is important to further narrow down the scope. Thus, the next step in the project is narrowing down the scope of the project before reframing the design question and requirements.

Key Conclusions

- *The ideal position to reach an appropriate level of trust is at the intersection between the use and pre-use phase*
- *Achieving the appropriate level of calibrated trust solely based on the HMI system of a vehicle is detrimental for the user as it creates an incomplete understanding of the system leading to poor predictability and reliability*
- *Communication is the key to achieving proper calibration of trust for the end user. However, the various types of communication (between internal stakeholders, company and user, user and AV) need to align together*
- *Impetuous to shift from a solutionism approach of design to a more reliable and realistic approach towards design such as socio-technical design*
- *Branding of the AV is as important as the technical design of the vehicle to achieve calibration of trust*
- *The introduction of AV leads to emergent behaviour form users.*
- *Transparency within the system can allow for developing AV with better capabilities of calibrating trust. But achieving transparency in itself is a challenging task*
- *A designer needs to take a global approach (overall vehicle) to the design of an AV not a local approach (autonomous system)*
- *The design of an AV is a process of improvement and not achieving perfection*
- *A designer needs to create a common language for communicating between the various stakeholders.*



"Trust is a very delicate area and needs to be built over time. Any one act can destroy faith which takes years to build"

Dr. Rachel Venturini

Photo by Museums Victoria on Unsplash

Part 2 I Chapter 4

New Design Question & Ideation

Within this chapter we create a new design question from the research insights gathered in the previous section. We then move onto explore the requirements set for the design intervention and the ideation process used in the project. We conclude with some of the initial concepts.

4.1 I Narrowing the Scope

Narrowing the scope of the design project was influenced by two factors: The fist was the *importance/ relevance of the problem space with regard to calibration of trust*. The second was the *duration of the project and the feasibility of the project under the guidelines of the pandemic*. Considering these factors certain problem spaces were eliminated like focus on designing for vulnerable road users as the process of calibration of trust for VRU is different from AV owners and designing solely for HMI systems.

From the analysis of the interviews and the literature review, it can be inferred that the product development process plays the most important role in designing for calibration of trust. The influence of the decisions made in this stage influences the pre-use phase as well as the use phase. For this reason it was concurred to focus on the product development phase. Finally, the relevance of the insights gathered were compared and two main areas were identified on which the final design should focus:

Socio technical approach to design: Through the literature and interviews it was seen that there is a need to focus on designing AV through a socio-technical approach of design. Considering the context of use and the social setting in which the vehicle will operate.

Communication: The design process of autonomous vehicles is a complex one with a number of stakeholders and thus communication plays a major role if we are to design an autonomous vehicle to calibrate trust. However, the communication is focused towards the internal stakeholders participating in the product development process and not the external stakeholders.

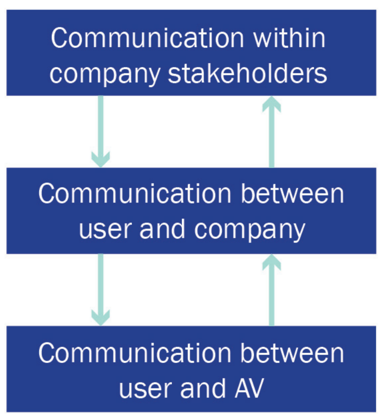


Figure 4.1. Three forms of communication that were identified during the literature review and interview with experts

4.2 I New Design Question

The initial design question considered was: *How to allow design teams of autonomous vehicles to design for human values (trust), for a multiple user context within urban locations ?* After going through the literature review and the data analysis phase certain the design question for the rest of the project can be rephrased as:

"How to empower the autonomous technology development team in the product development process to design for a socio-technical context resulting in calibration of trust (understandability and predictability) in the pre-use and use phase between the user and their autonomous vehicle?"

The design question was chosen through the process of co-evolution of the problem space and the solution space (Dorst & Cross, 2001). This was performed to achieve a design direction that fit with the insights gathered and at the same time build on the existing research and design of calibration of trust. The design question focuses on the aspect of understandability and predictability, leaving out reliability. The reason for this reliability is created by using the autonomous system multiple times leading to a belief that the system will work. This part of calibration of trust falls in the use phase and cannot be designed into the product development process.

Focusing on the phrase "empower designers/developers" from the design question to dive deeper into what are the sub parts of the design question. From this process there came out three parts that are important to designers/developers. These three sub-parts were selected based on the insight gathered during the literature review and interviews conducted with experts:

Empowering designers to inform/sensitize the design and development team about trust and calibrated trust

Approach the development of autonomous functions in a socio-technical context

Create clear communication of autonomous capabilities across multiple stakeholders(internal and external)

4.3 | Design Requirements

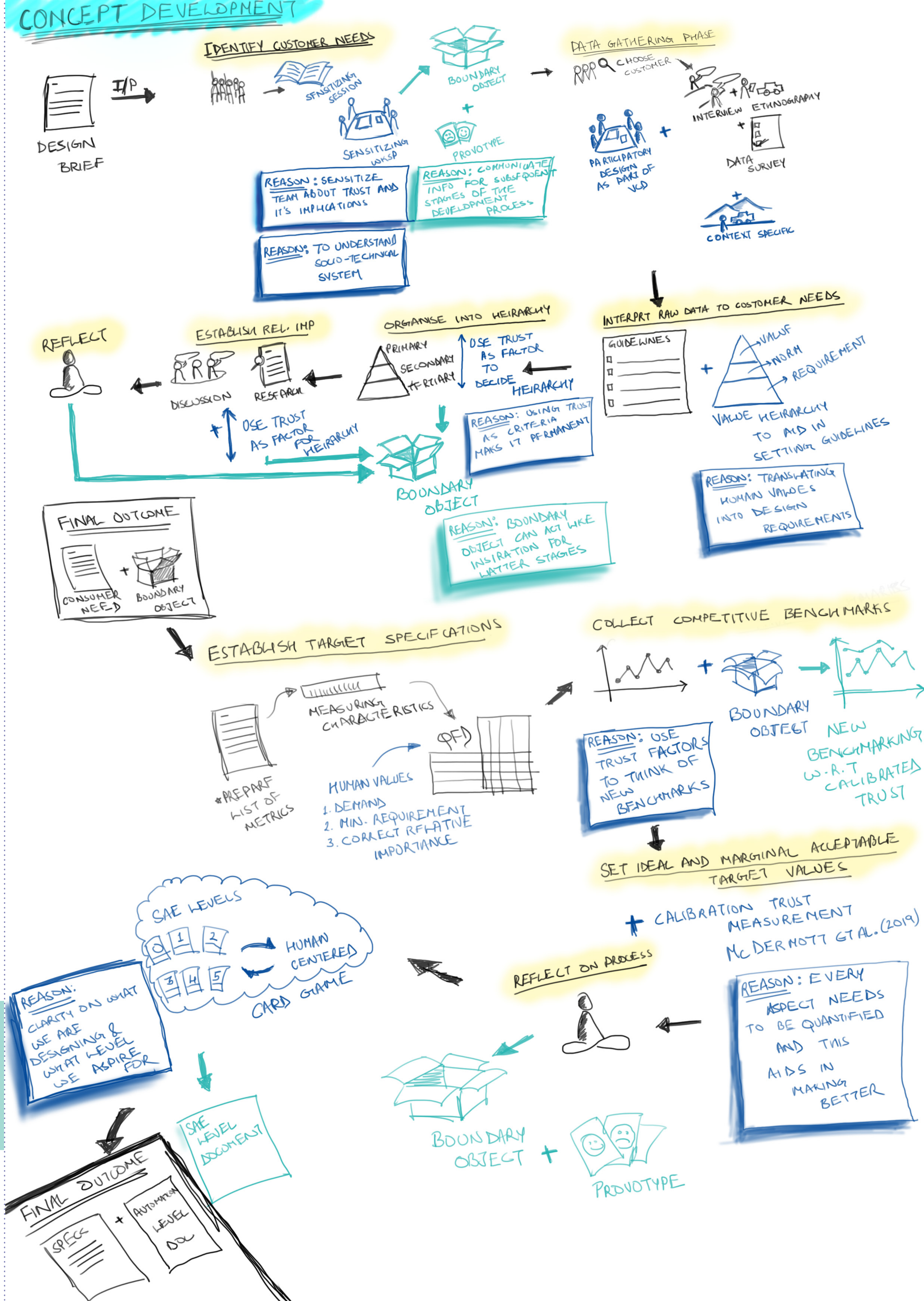
- The design must create understandability of the autonomous system among the various stakeholders
- The design must assist in evaluating predictability of the system for the autonomous technology development team
- The design must integrate within existing product development process used in the automobile industry
- The outcome must provide a holistic overview or an autonomous function and not a technical overview (solutionism approach)
- The outcome must provide developers an understanding of the context of use of autonomous vehicles when designing the individual functions
- The design must provide an understanding of trust and the factors that influence the formation and calibration of trust
- The design must allow for conveying the understanding of trust across multiple stakeholders throughout the design process
- The design must allow for further modifications to accommodate for changes in the design process of autonomous vehicles
- The outcomes of the design should integrate into the design of HMI systems of the autonomous vehicles
- It should have an intuitive way of learning which is easy and fun at the same time.

4.4 | Ideation Process

The process of ideation followed a process similar to co-evolution, in which the design question and the solution space both evolved until a satisfactory solution space was found (Dorst & Cross, 2001). This was necessary as the product development process for automobiles is complex (Ulrick et al., 2019) and creating a design intervention at each stage would make the project more complex and difficult to complete in the stipulated amount of time. The process of co-evolution used in the project has been detailed below:

- *Reflecting on the product development process from the literature and comparing the model proposed by Ulrich et al.,(2019) with the development process present in the automotive industry.*
- *Break down the product development process into its constituent stages and sub-stages. Followed by describing the input and output of each stage along with the desired goal of the stage.*
- *After creation of the sub-stages, and detailing them out. Concepts were generated within each stage that would satisfy the proposed design question.*
- *Having generated a number of concepts at individual sub-stages of the process development process, I zoomed out to look at the complete process. This allowed us to look at the various concepts as a whole picture.*
- *The concepts generated were evaluated by considering the requirements, in some cases concepts were combined to form new concepts.*
- *The above process was performed in an iterative manner and in the end four design interventions were created that met the overall design question and the sub-design questions that were created.*

Figure 4.2 : Shows a collection of the initial concepts that were generated using the ideation process as described above. The black portion represent the conventional product design process parts. The blue portion are the design intervention. The green represents the outcome of the design intervention . The figure represents the ideation within the concept development phase of the product development process.



Part 2 I Chapter 5

Calibrated Trust Toolkit

The final outcome of the project is explained within this chapter. Beginning with an overview of the toolkit, we further dive into the individual parts of the toolkit. Exploring the iterations, use case and position of the toolkit within the product development process

5.1 I Introduction

Toolkit & Trust

The Calibrated Trust Toolkit is built on four design interventions that are designed to collectively aid in designing for calibration of trust. The four design interventions address the three sub-parts of the design question: I*nform about calibration of trust, designing within a socio-technical context and communicating about trust*. The design interventions are as follows:

- **Sensitizing session**: A session during the planning phase that allows for the design and development team to create a common understanding of trust and the factors influencing them. This is followed by creating a trust goal that is in line with the design direction.
- **Autonomous Function Visualizer**: A human centred visualization canvas that allows developers to break down individual autonomous functions into technology, context of use and user behaviour. This creates a more holistic understanding of the autonomous, easy communication across various functions and the ability to capture iterations performed on the autonomous function throughout the development process.
- **User Decision Matrix**: The user decision matrix is used in the detail design process to design HMI and eHMI systems. The basis of the matrix is based on how we make decisions, and what role the context of use plays in the making decisions.
- **Trust Enhancing Communication (TEC)**: TEC is a set of principles and requirements that are created for communicating with external stakeholders when moving from the product development process to the pre-use phase.

The design interventions were aligned with the trust model as proposed by Hoff & Bashir (2015) as seen in Figure 5.2. Trust has been depicted as a pyramid with dispositional trust as the base over which situational trust is placed and at the top there is learnt trust. The reason for placing dispositional trust at the base of the pyramid is because it is the collective term used to define trust characteristics that are innate to the user such as the influence of culture, their age, gender and personality (Hoff & Bashir, 2015). These factors cannot be influenced by design but on the contrary need to be considered during the design process. This is the reason for creating a sensitizing session, encouraging designers to consider factors such as culture, personality during the initial research and benchmarking of specifications.

The second layer of the pyramid is situational trust, which represents the part of trust that is influenced by the situation or context (This includes internal context such as stress, exhaustion as well as external situational factors such as the weather). The user decision matrix and the autonomous function visualiser are designed to address this part of trust.

The last part of the pyramid is the learnt trust. This part is the information and understanding we have with regard to an autonomous vehicle or an autonomous function. The Trust Enhancing Communication and the Autonomous Function Visualizer fall in this category. We see that learnt trust spans not just the product development process but the pre-use phase , this is because for the end user receives most of the information about an autonomous vehicle during the pre-use phase and in the use phase. However, this information is a product of the product development process and the pre-use phase

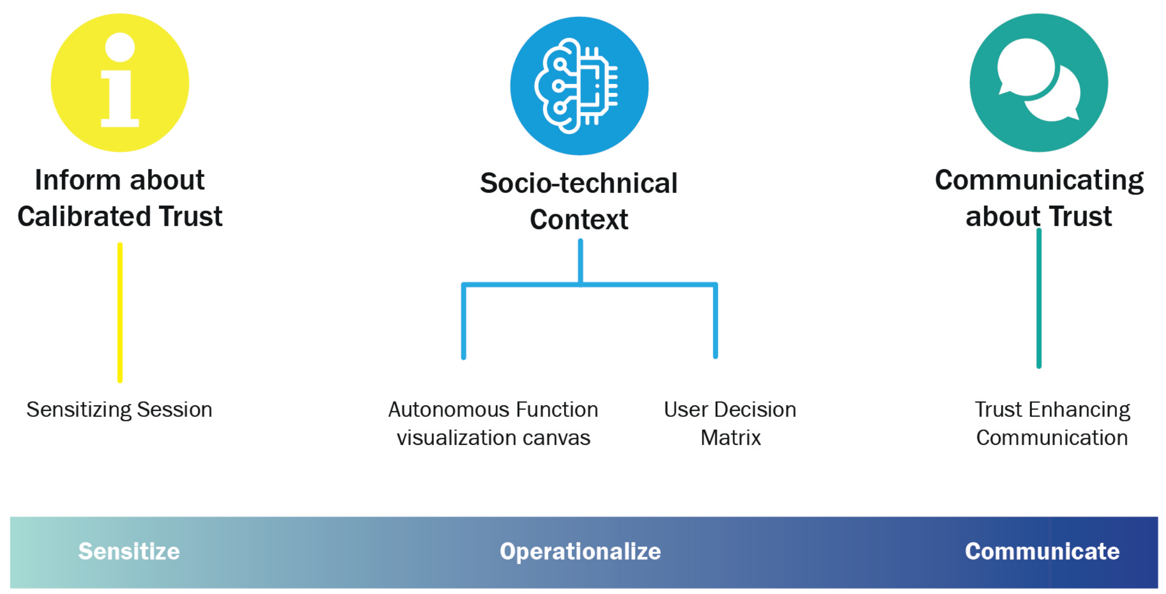


Figure 5.1. Breakdown of the toolkit with regard to the three main aspects of the design question

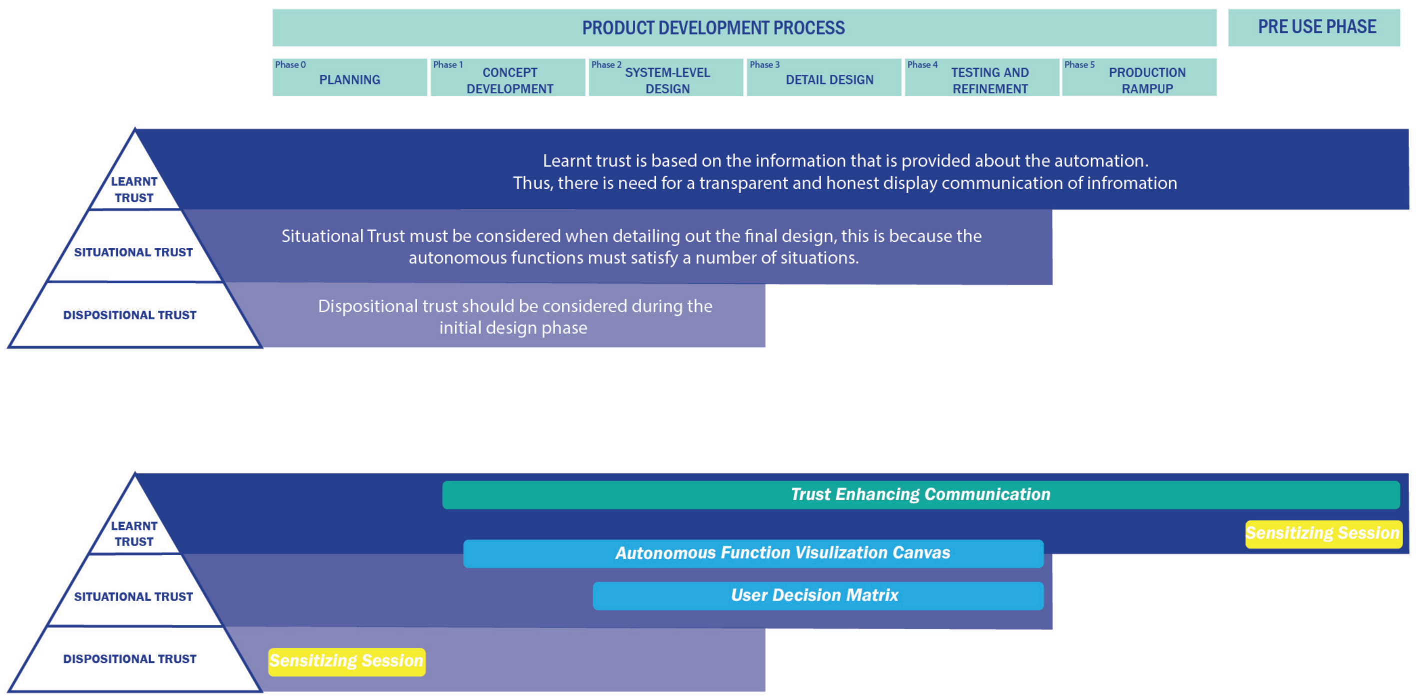


Figure 5.2. The relation of the various parts of the toolkit with trust, when viewed across the product development process. The top figure shows the influence of the various types of trust and the various phases of the product development process. The bottom figure highlights the relationship of the four parts of the toolkit to the various types of trust

Toolkit & Calibrated Trust

We have established relationships between the various types of trust and the relationship with the toolkit. We now explore how the toolkit aids in calibration of trust. As discussed in Part-1 of the project, calibration of trust is the *active management of understandability, predictability and reliability*. The toolkit focuses on understandability and predictability. Reliability is not considered, as its importance is in the use phase and the testing stage of the product development process.

The sensitizing session addresses both understandability and predictability, by first diving into the reasons for distrust or trust and then looking forward to the consequences i.e. predictability. The autonomous function visualization canvas is purely dedicated towards the understandability of the system while the user decision matrix is focused towards developing an understanding of user behaviour which falls under predictability. The Trust Enhancing Communication like the sensitizing session is developed to create both understandability and predictability within the stakeholders.

RECAP

Dispositional Trust: *represents an individual's overall tendency to trust automation, independent of context or a specific system (Hoff & Bashir, 2015). Factors such as culture, age, gender and personality traits influence is considered in this type of trust.*

Situational Trust: *It represents the context in which the automation is used. Situational trust can be further broken down into external variability, that includes variables such as workload, task difficulty, type of system and the system complexity. Internal variability which includes factors such as self-confidence, mood and attention. This is more focused on the current state of the user when interacting with the automation*

Learned Trust: *The final part of trust development is learnt trust which is a representation of a users evaluation of a current system drawn from the past experience or the current situation (Hoff & Bashir, 2015). Learned trust is further divided into two sub parts:initial learned trust and dynamic learned trust.*

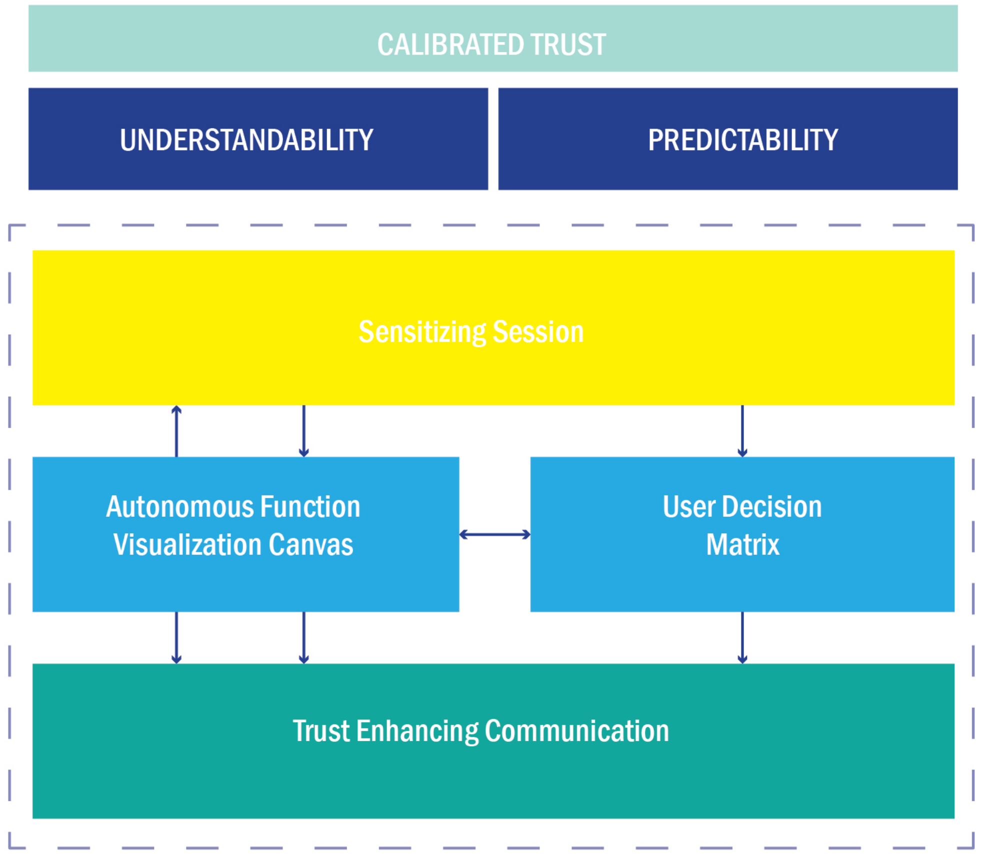


Figure 5.3 The relation of the parts of the toolkit with calibrated trust (understandability and predictability). The dashed box represent the boundaries of the toolkit. The arrows between the individual parts represents their interrelation with individual parts of the toolkit.

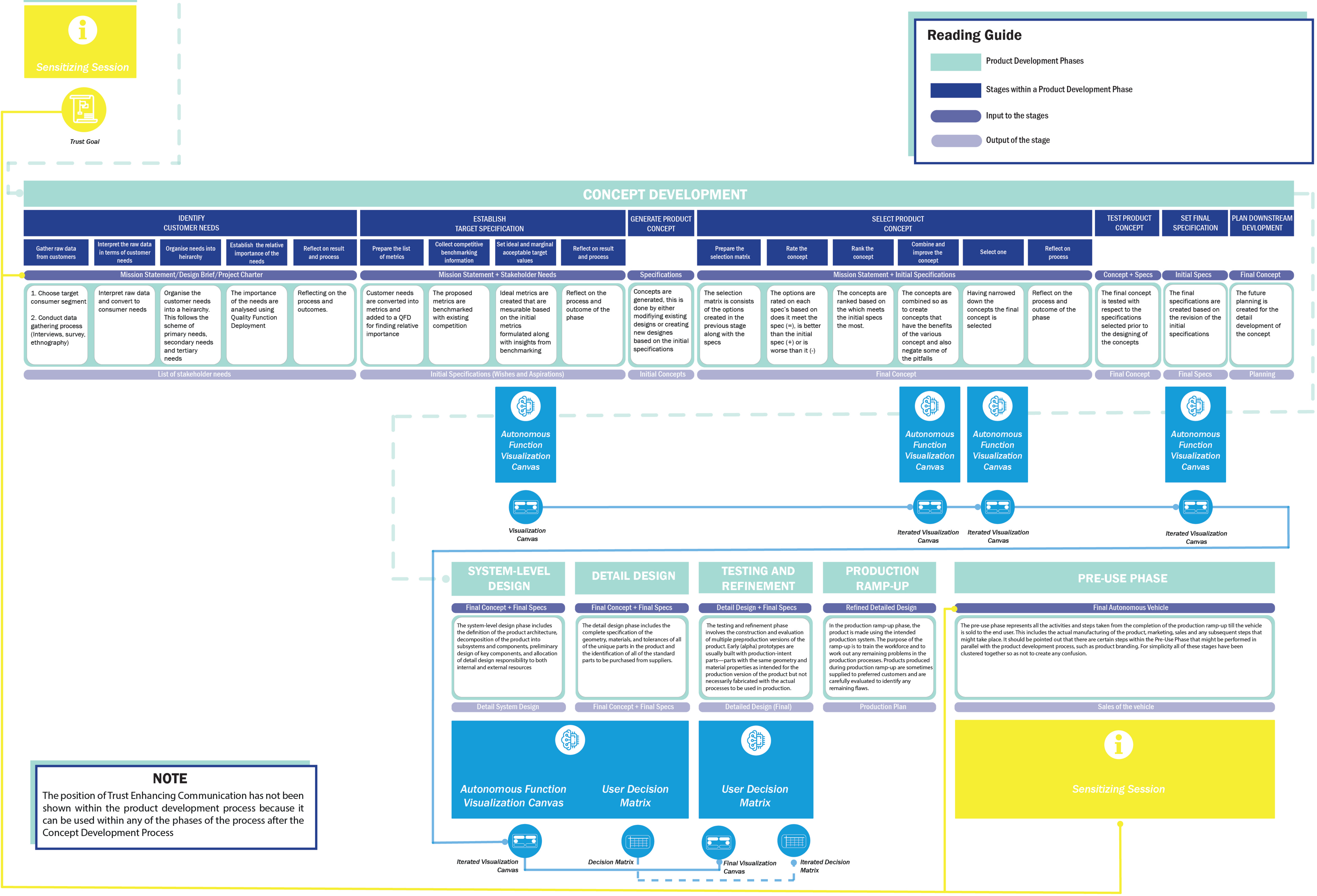
Positioning the toolkit with the product development process

The calibrated trust toolkit is positioned in multiple phases of the product development phase. This is done to meet the iterative nature of the product development process and the complexity of the process (Ulrick et al., 2019) because of the number of the systems present within an automobile (Bhise, 2017). The detailed positioning of the individual aspects of the toolkit can be seen figure on the next page. More details with regard to the positioning of the individual parts of the toolkit can be found in the individual sections in the subsequent section.

The sensitizing session takes place in the planning stage of the product development process, this is because it allows for conducting the sensitizing of all stakeholders (primary and secondary) stakeholders before the design process of the vehicle begins. The autonomous function visualization canvas is first introduced in the concept development stage, under the "*Establish Target Specifications*". This is further iterated upon in the "*Selecting Product Concept*", "*Set Final Specifications*". Moving to the *System-Level Design* and *Detail Design* Phases the autonomous function visualization canvas is further iterated upon and we create the final visualization canvas at the *Test & Refinement* phase of the product development process.

The user decision matrix is used within three phases of the development process, the first is the *System Level Design* followed by the *Detail Design* and the testing phase. The reason for the use of the toolkit in multiple phases of the design process is because of the various scenarios in which it can be deployed (discussed in detail in the section on user decision matrix). The Trust Enhancing Communication is positioned to be used within the *Concept Development Process* and subsequent phases of the product design process.

Calibrated Trust Toolkit positioning within the Product Development Process



Sensitizing Session

5.2 | Sensitizing Session

The sensitizing session focus is to create an understanding of calibrated trust and the underlying factors that influence the creation and the maintenance of calibrated trust. The session is based on the concept of generative design. In which the participants follow the say, do and make process of expressing their thoughts (Sanders & Stappers, 2012). The aim of the sensitizing is to reflect on personal understanding of trust and learn about calibration of trust. The sensitizing session is placed as the first design intervention in the planning phase to sensitize the design and development team before they begin to use the other tools. This is important because we want the team to think about the impact their design has in terms of users' trust and the implications poor design or unintended design has on the users' understanding of the product. The sensitizing session is based on the insight provided during the expert interviews that it is important to have a holistic picture of how trust is built between the vehicle and the user. In addition, the sensitizing session acts as the foundation for calibration of trust and the subsequent parts of the toolkit.

The outcome of the sensitizing session is a trust goal that the team can associate with and pursue throughout the design and development process. This goal is the consolidation of the discussion during the sensitizing session. The goals allow communication of the outcomes to other internal stakeholders who were not present in the sensitizing session. At the same time is a reminder once the participants dive deeper into the product development process.

The sensitizing session is divided into three stages. The stages are inspired to be in line with the generative design process (Sanders & Stappers, 2012) and also incorporate the idea of reflection in action and reflection on action (Schon, 1983). The stages are as follows:

The sensitizing booklet/package: The sensitizing package is provided to the participant before the sensitizing session. This consists of a booklet (digital/physical), the participants are requested to complete the package before the main session so that it can be discussed during the session.

Sensitizing session: The main sensitizing session is to have a discussion on the topic of calibrated trust and the factors that influence it. The session starts by reflecting on the sensitizing booklet and building on the content of the sensitizing package.

Reflection package: At the end of the session the participants are provided a reflection package which allows them to reflect back on the session and write down key insights they have found. This is a package that the participants are free to take back with them.

The Facilitator

Conducting a sensitizing season requires certain expertise and experience. For this reason one member of the team must be appointed as the "Facilitator". The responsibility of the facilitator is to schedule and conduct the session. This not only includes the logistics of the session, for example setting up the room for the sensitizing session, acquiring all the material that will be consumed during the session. But conducting the session and directing the session. Thus a facilitator should either be or have experience as a Scrum Master, Agile Coach or Creative Facilitator. In addition basic knowledge in the field of trust between human automation would be an added bonus. The responsibilities of the facilitator have been described below:

Pre-Session

- *Planning the sensitizing session, along with securing physical locations for conducting the sessions*
- *Selecting the participants for the sensitizing session, contacting them and scheduling the date of the session*
- *Creating the sensitizing packages and delivering them to the participants (In certain cases the participant might have question, the facilitator is required to address these queries)*
- *Preparing the material for the sensitizing session*

During the Session

- *Facilitate the session and discussion*
- *Record the discussion on the canvas*
- *Maintain the schedule of the session, including bio-breaks and refreshment*
- *Converging at a Trust Goal*
- *Hand participants the reflection package*

After the Session

- *Collect the canvas and sensitizing package to store or discard*
- *Communicate the Trust Goal to the other team members along with a brief explanation*

	Sensitizing Package	Sensitizing Session	Reflection Package
AIM	Reflect on personal belief with respect to trust and translate those insights into trust with respect to an autonomus function	Develop collective understanding of trust and its importance when using an autonomous function	Record personal observations and insights for reflecting or referring later in the product development process
OUTCOME	Explore despositional trust and personal understanding of trust	Create collective understanding of trust and understand calibration of trust	Reflect on the session and the results of the session
DURATION	30 minutes (Spanned over a week)	4.5 Hours (Single Session)	20 minutes (Single Session)
TYPE	Individual Activity	Group Activity	Individual Activity
RESOURCES	Print out of Sensitizing Booklet(x number of participants) Camera (Depending on the selected autonomous function)	Print out of Sensitizing Canvas Post-It Notes Markers Sticky Dots	Print out of Reflection Booklet(x number of participants) Markers

Table 5.1. A detailed overview of the three parts of the sensitizing session

Participant Selection

Selection of the right participants is important during the process of sensitizing. The mixture of participants must three categories:

Expertise across multiple functions: Designing a vehicle is considered a complex design process with a number of functions and sub-functions such as power transmission, steering and suspension, aerodynamics, ergonomics, braking, etc. Thus, a collection of participants must be chosen that can cover a majority of expertise.

Corporate Ladder: The second factor is to gather stakeholders from various levels of the corporate ladder to take part. Which would include engineers/developers to function heads and even department heads taking part.

Influence Factor: The influence factor considers the association the participant will have with the development process of the vehicle, in terms of responsibility within the development process, the time dedicated to the project and the ownership of the project or sub-parts of the project.

The need for this diversity of participants is important to achieve a homogeneous representation of the various stakeholders that will take part in the design and development of the autonomous vehicle. The session should have around 5-7 participants. This number allows for a diverse representation of stakeholders but at the same time a manageable group of participants to conduct the session. If there are an excessive number of participants, the number of sensitizing sessions can be increased and the intensity of the session decreased. This again depends on the time and availability of resources.

NOTE: *Detailed description of the role of the facilitator and other information with regard to participant selection and conducting the session can be found in the Calibrated trust toolkit manual*

Figure 5.4 Mock-up of the sensitizing package

Sensitizing Package

The aim of the sensitizing package is to provide an initial insight into the participants belief and understanding about trust. The sensitizing booklet is to be provided to the participants a week before the actual session. It requires the participant to fill in a page each day and should not take over 10-15 minutes per day to finish the content of the page. It is important to note that the sensitizing package should not be cumbersome to use or have excessive activities to complete as it will deter the participant from completing the package. Another important point to keep in mind is that it must not hinder the day to day activities of the participant.

The sensitizing package covers two major themes within it: *dispositional trust* and *trust towards a specific autonomous function*. The reason for selecting the theme of dispositional trust is because it is the collective term given to the individual characteristics of an individual. The sensitizing package allows for thus reflection on these characteristics that are part of dispositional trust. Trust in a specific automation narrows down the scope to a particular function. It should be noted that this function will be used in the main session as well so the care must be taken that the participants have all interacted with the described autonomous function.

Sensitizing Package

Aim : Record personal belief and understanding about trust. Introduce the participant to the trust between human and automation

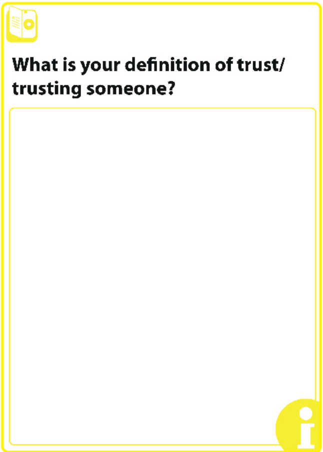
Duration : 30 minutes (Spread over a week)

Type of Activity : Individual Activity

Material Required : Copy of the sensitizing package (digital/physical), a camera to take images depending on the autonomous function selected

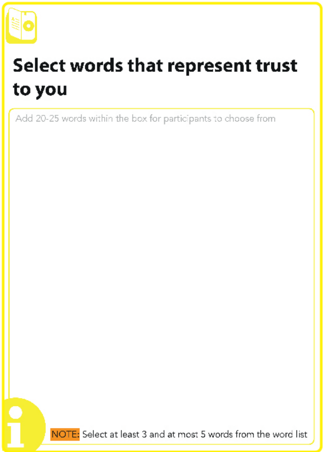
Sensitizing Package Breakdown

1. Dispositional Trust



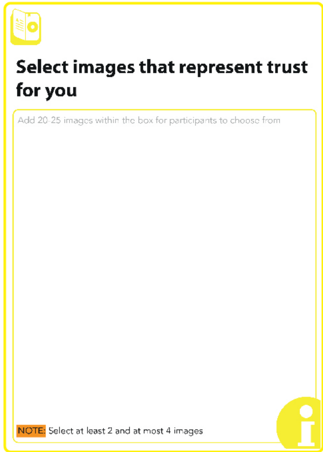
1.1

Participants are requested to describe the meaning of trust, they are free to even use images or photographs to express their answer



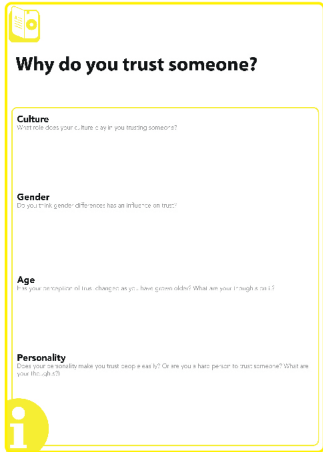
1.2

Select words and phrases that represent trust from a list of words



1.3

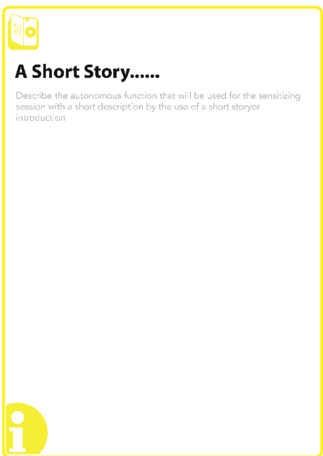
Similar to the previous step selecting images that represent the participants idea of trust



1.4

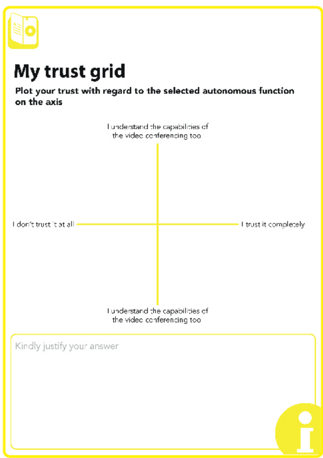
Participants are encouraged to comment on how the four factors of dispositional trust influence their own trust

2. Trust in Automation



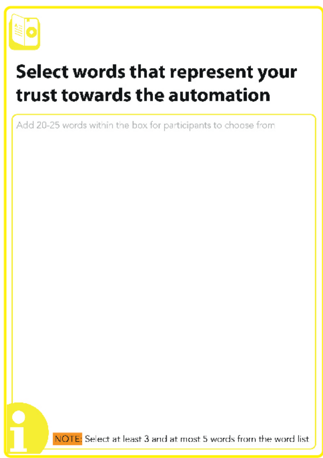
2.1

Introduce the automation that will be used in the sensitizing session by formulating a short story explaining the use of the automation



2.2

The participant plots their trust level along the axis and explain their reasoning



2.3

Participant select words from a word list that represent their trust towards a specific automation

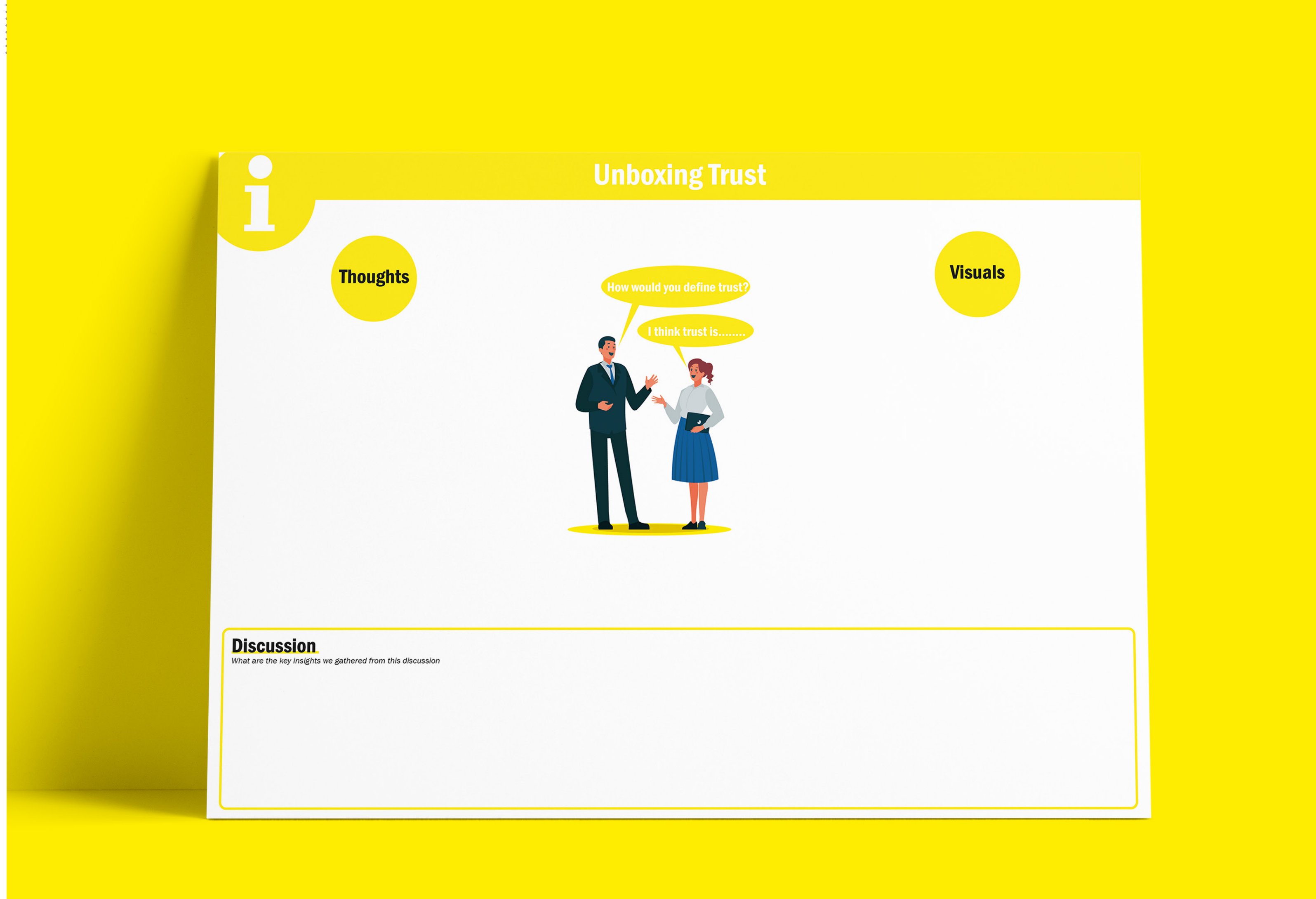


Figure 5.5. Mock-up of one of the seven canvases used during the sensitizing session

Sensitizing Session

The sensitizing session builds on the insights gathered with the sensitizing package. The aim of the sensitizing session is to create a common understanding of the factors influencing trust, resulting in creation of a trust goal. The sensitizing session consists of seven canvases, these canvases are used for directing the discussion during the session and recording insights. The canvases are designed to follow the same order as that of the sensitizing booklet initially and then add on to them by discussing the concept of calibrated trust, the process of trust formation and certain trust misconceptions.

The canvases have been designed in a manner that only one specific activity is performed using them. This activity is done in the top half of the canvas. At the bottom part of the canvas a discussion section is created. This is done so that key insights and points can be readily recorded in the section allowing the participants to revisit canvases when brainstorming about the Trust Goal and not get lost in the post-its or other information jotted on the canvas.

While seven canvases might seem to be a lot of canvases to go through in one session, all canvases are interrelated and are designed to facilitate the direction of the discussion along with providing a format of recording the insights and data. In general the group must not spend more than 30 minutes on a single canvas. Which means that the overall time required for conducting a session would be 4.5 hours (including breaks).

Sensitizing Session

Aim : Build collective understanding of calibrated trust and reach a common trust goal

Duration : 4.5 hours (Single session)

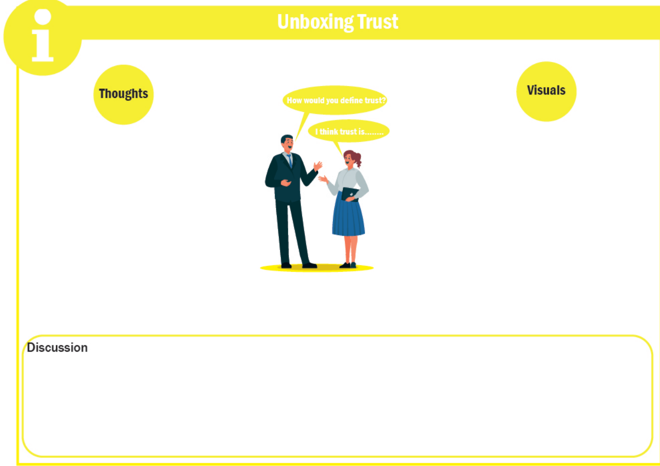
Type of Activity : Group Activity

Material Required : Print out of the canvases, Post-its, markers, dot voting Post-its, masking tape

Overview of the Canvases used in the Sensitizing Session

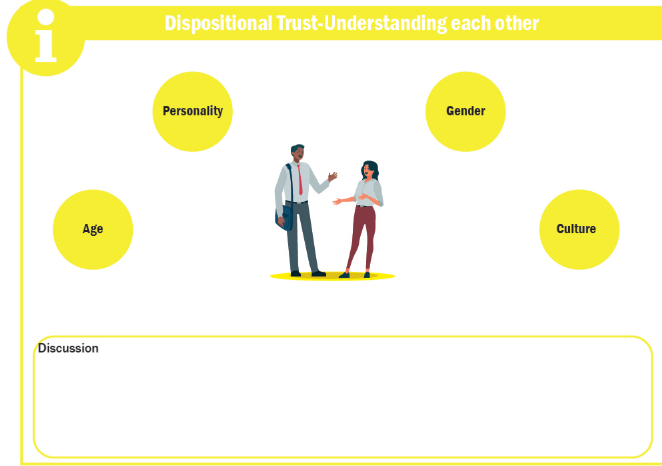
Canvas-0
“Welcome!”
20 minutes

Introduce the plan of the session and set some ground rules



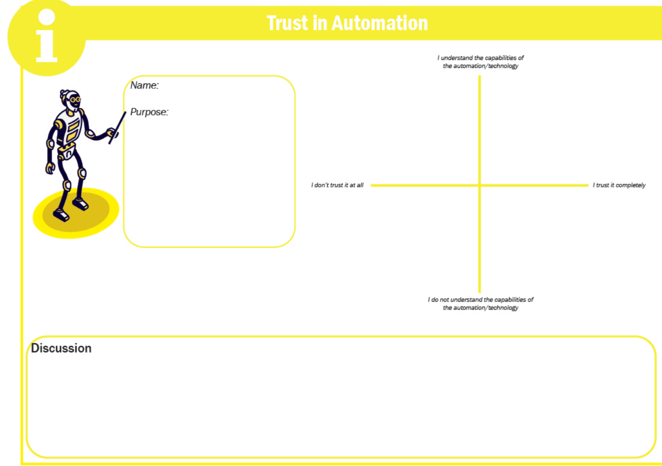
Canvas-1
“Unboxing Trust”
30 minutes

Explore the groups understanding of trust in general, using the sensitizing booklet as a starting point



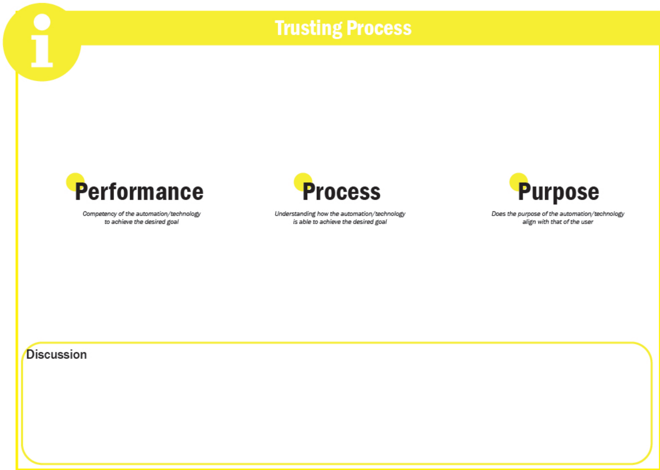
Canvas-2
“Dispositional Trust”
20 minutes

Dive into the four factors that influence dispositional trust and understand the groups experience with respect to the four factors



Canvas-3
“Trust in Automation”
30 minutes

The canvas collects the insights about trust in automation as discussed in the sensitizing booklet. Also introducing calibrated trust

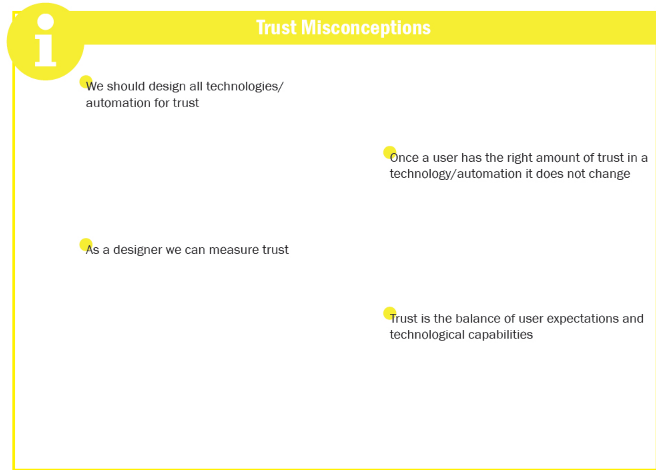


Canvas-4
“Trusting Process”
30 minutes

Explore the process of trust/ distrust with respect to the automation by breaking it into the trust formation process

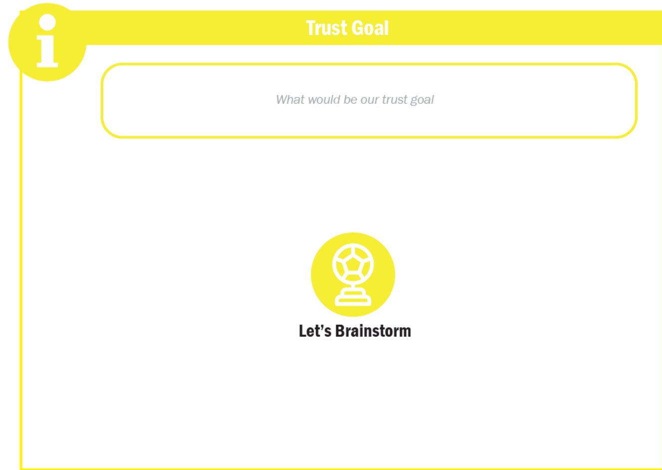
Canvas-5
“The Devils Advocate”
30 minutes

Use negative analogies to explore situations that can create distrust



Canvas-6
“Trust Misconceptions”
20 minutes

Discussion on certain misconceptions we have with respect to trust



Canvas-7
“Trust Goal”
30 minutes

Brainstorm on creating a trust goal that will be the outcome of the session

Figure 5.6. Mock-up of the reflection package

Reflection Package

The reflection package is for participants to record their personal reflections from the session and translate them to their role in the product development process. It is provided to the participants before the sensitizing session starts so that they can make personal notes in one document. The aim of providing a reflection package in addition to creating a trust goal is to aid in the capturing of insights and perspectives throughout the sensitizing session and at the same time allow for reflecting back once the session is over and they return back to their respective functions. While the outcome of the sensitizing session will be recorded and shared with the participants by the facilitator. The reflection package is more for recording personal insights.

The reflection package is divided into two parts: personal reflection and professional reflection. Within the personal reflection the participant can fill in insights that do not directly relate to their function and role in the development of the autonomous vehicle. The professional reflection part is directed addresses the question of "*How does my function/ role in the product development influence the trust goal?*".

Sensitizing Package

Aim : Record personal reflections based on the discussion and outcome of the sensitizing session

Duration : 20 minutes (After sensitizing session)

Type of Activity : Individual Activity

Material Required : Copy of the reflection package (digital/physical)

Positioning in Product Development Process

The sensitizing session is designed to be conducted in the planning phase of the product development process. The planning phase is also known as Phase-0 as it precedes the project approval and the product development process (Ulrick et al., 2019). This makes it the ideal position for sensitizing the product development team, so that the insights of the sensitizing session can also be incorporated within the planning of the product development process.

A second possible position to perform the sensitizing session is the pre-use phase. This session is conducted so that the tertiary stakeholders (dealership, tier-1 suppliers, marketing department etc.) can also be brought on the same understanding of the importance of trust and sensitized about the importance of calibration of trust. There however does exist a small difference in the sensitizing session in both cases. In the sensitizing session conducted at the Planning Phase we end with a trust goal, in case of the session conducted in the Pre-Use Phase we do not end with a trust goal instead we reflect on the trust goal as formulated in the first session. This is done so that there are not two separate goals for the same vehicle. Thus, in the last canvas of the sensitizing session instead of creating a trust goal the participants use the reflection package to explore the implications of the trust goal towards their responsibilities.

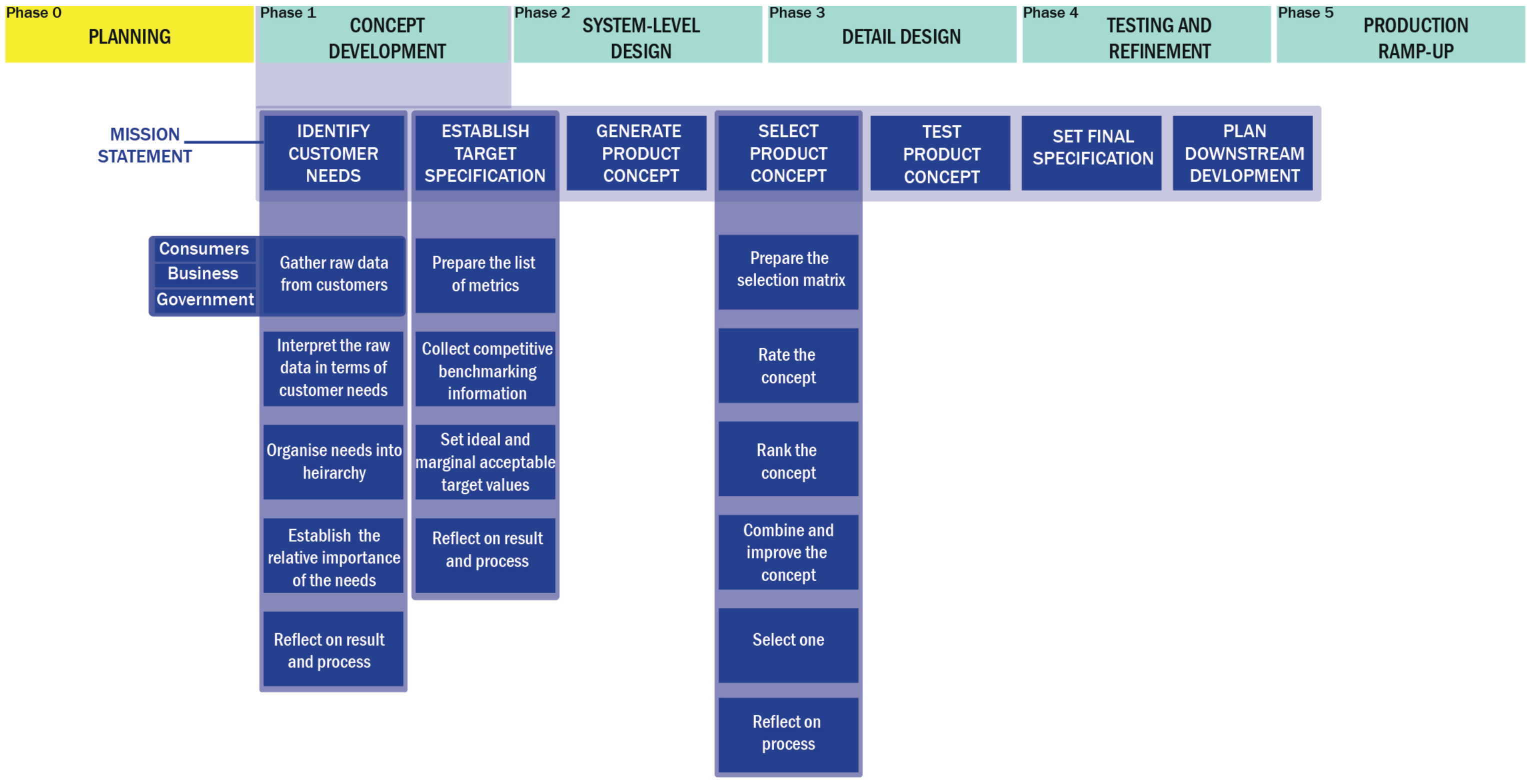


Figure 5.7. The position of the Sensitizing session within the product development process

Autonomous Function Visualization Canvas

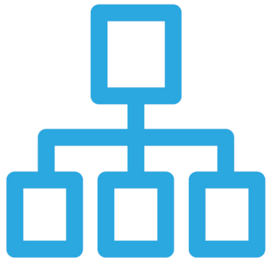
5.3 | Autonomous Function Visualization Canvas

The "Autonomous Function Visualization Canvas" is based on the concept of moving away from a technical based description of the autonomous functions to a human centered approach of visualizing the functions. This is achieved by breaking the function into its constituent parts starting from the technology to the context and finally the users actions. The aim of the design intervention is to always keep the user's actions and context of use in mind when describing the autonomous function. Further, the canvas can be used to visualize the overall automation function in one connected document. The design intervention builds on the research of Yang et al.(2017), in which the 6 levels of automation proposed by SAE are compared to 5 levels of human centered levels of automation (HuLOA). While the research does provide a more human centered understanding of the SAE levels, there exist limitations. Car manufacturers rarely describe autonomous functions corresponding to the SAE levels. In addition the HuLOA, only highlights the desired actions of the driver/user of the vehicle. These two shortcomings are addressed in the Autonomous Function Visualization Canvas.

The second use of the visualization canvas is in the form of a semantic boundary object (Carlile, 2002) i.e. it creates a common understanding of autonomous function within the product development process and pre-use phase. The visualization canvas must not be considered as a constant document but an iterative document that is updated as we move along the product development process.

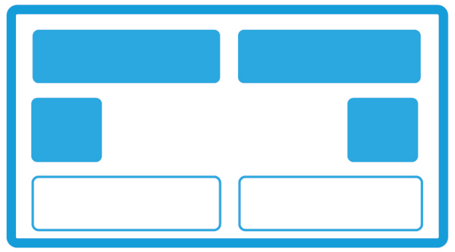
The autonomous function visualization canvas is designed for the autonomous technology team, unlike the sensitizing session that should be used by the combined product development team. It is also the first of two parts of the toolkit that are focused on the operationalizing of calibrated trust within a socio-technical context, the other being the user decision matrix.

The Autonomous Function Visualization Canvas consists of two parts. The first is the Function Tree and the second is the canvas itself. The Function Tree is a visualization of all the individual autonomous functions and combined functions(using multiple autonomous functions at the same time). It is important to segregate individual functions from combined functions because they might give the wrong impression to the end user. Considering the example highlighted by Mirnig et al.(2016), some SAE level 2 vehicles have the capability to use adaptive cruise control and lane assistance simultaneously, this creates the illusion of driving a level 3 vehicle but in actuality the combination of adaptive cruise control and lane assistance will not work in all conditions that a level 3 vehicle should function. The second part of the Autonomous Function Visualization Canvas is the visualizing canvas which breaks down the selected autonomous function to create a more human centered understanding of autonomous functions.



Function Tree

Overview of the diagram that shows the connection of the various functions and how they relate to each other. This is the predecessor of the canvas and allows to get an overview of the complete automation and then focus on one single function.



Visualizing Canvas

The visualization canvas consists of four parts. The first part is the function description followed by the technology used. The last two parts consists of the context of use and the ideal user behavior.

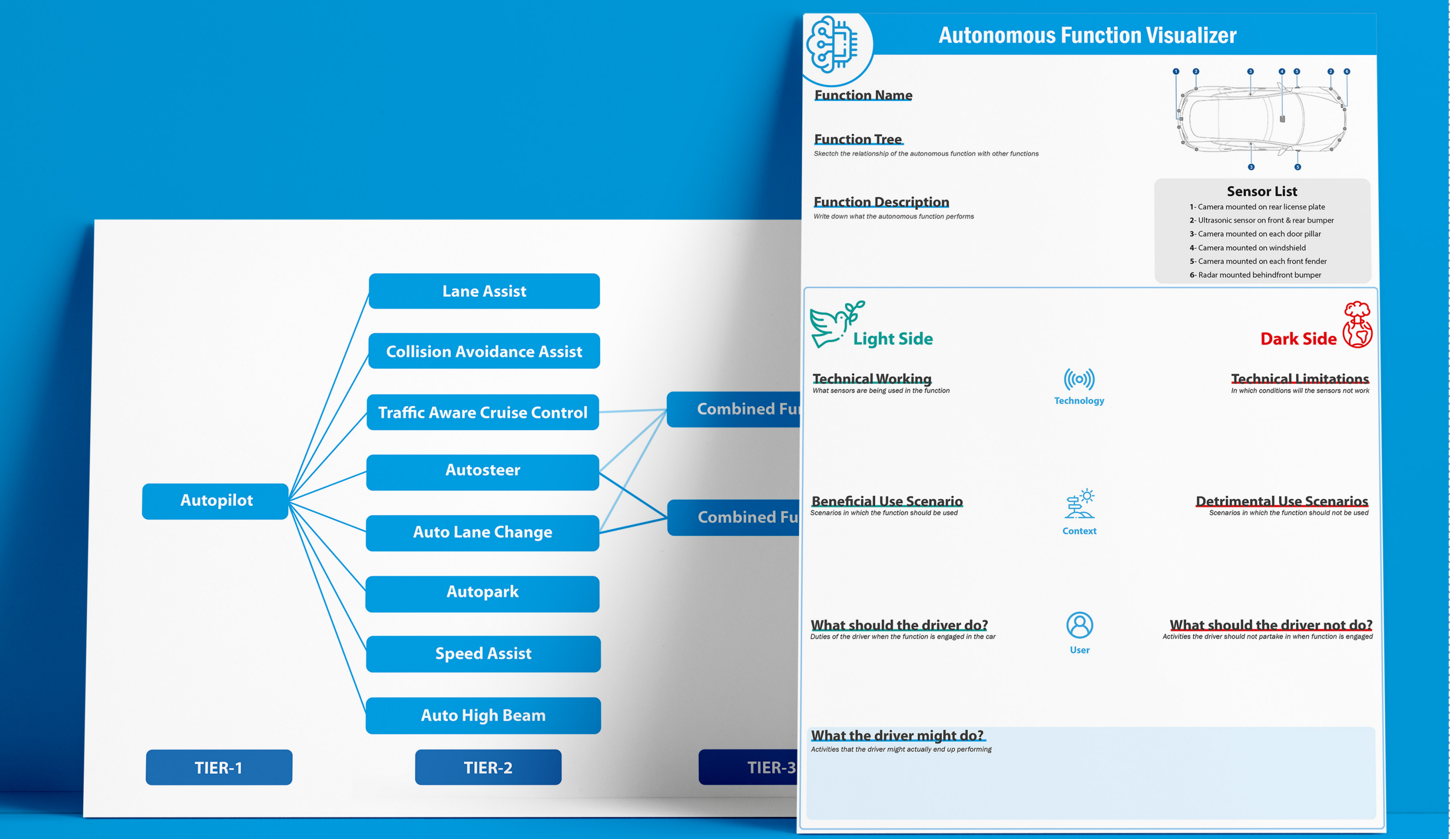


Figure 5.8. Mock-up of the autonomous function visualization canvas and the function tree

Function Tree

The function tree as described previously is a visual representation of the various autonomous functions present within an autonomous vehicle. The function tree is divided into 3 tiers, the first tier is the generic name given to the overall automation present within the vehicle (example is Autopilot, the name given to the autonomous system within Tesla vehicles). Tier two is the list of individual automations present within the vehicle. The identification of an individual function can be challenging as there are certain overlaps that exist between various functions. For this reaon the "sense-think-act" approach of defining and distinguishing a function is employed. The third tier is combined functions that use more than one function simultaneously. The prominence of a third tier autonomous function is more prominent in level-2 and 3 automation, as we move towards higher levels of automation, tier-2 and tier-3 will merge. However, it is difficult to predict how this would look like, as we have not reached higher levels of automation. An example of the function tree has been highlighted below in *Figure 5.10.*

Defining a Function

An autonomous function is defined on the basis of the "sense-think-act" loop. Each function must differ from the others in one of the three aspects to be segregated into a separate autonomous function (Haddal et al., 2018):

Sense: *The information it is receiving from the environment through sensors*

Think: *Process the data according to the required software*

Act: *Perform the required action based on the outcome of the previous step*

To illustrate the process of defining an autonomous function consider the example of Lane Assistance and Autosteer. From the table we can see that there are overlaps in the Sense and Think steps of the loop, however there is a difference in the Act step. This shows that there is a need to segregate both these autonomous functions and not cluster them together.

	Sense	Think	Act
Lane Assistance	Uses cameras and ultrasonic sensors to monitor lane markings as well as surrounding areas	Calculates the distance and location of the vehicle with respect to road markers and other road users	Displays the information to the driver to make a decision on future actions
Autosteer	Use cameras, ultrasonic sensors and radar sensors to detect lane markings and the presence of vehicles and objects	Calculates the distance and location of the vehicle with respect to road markers and other road users	Maneuvers the vehicle to keep it within the lane markings and also performs lance change when the turn signal is activated

Table 5.2. Using the sense-think-act loop to distinguish between the Lane Assitance and Autosteer autonomous functions present within a Tesla Model S

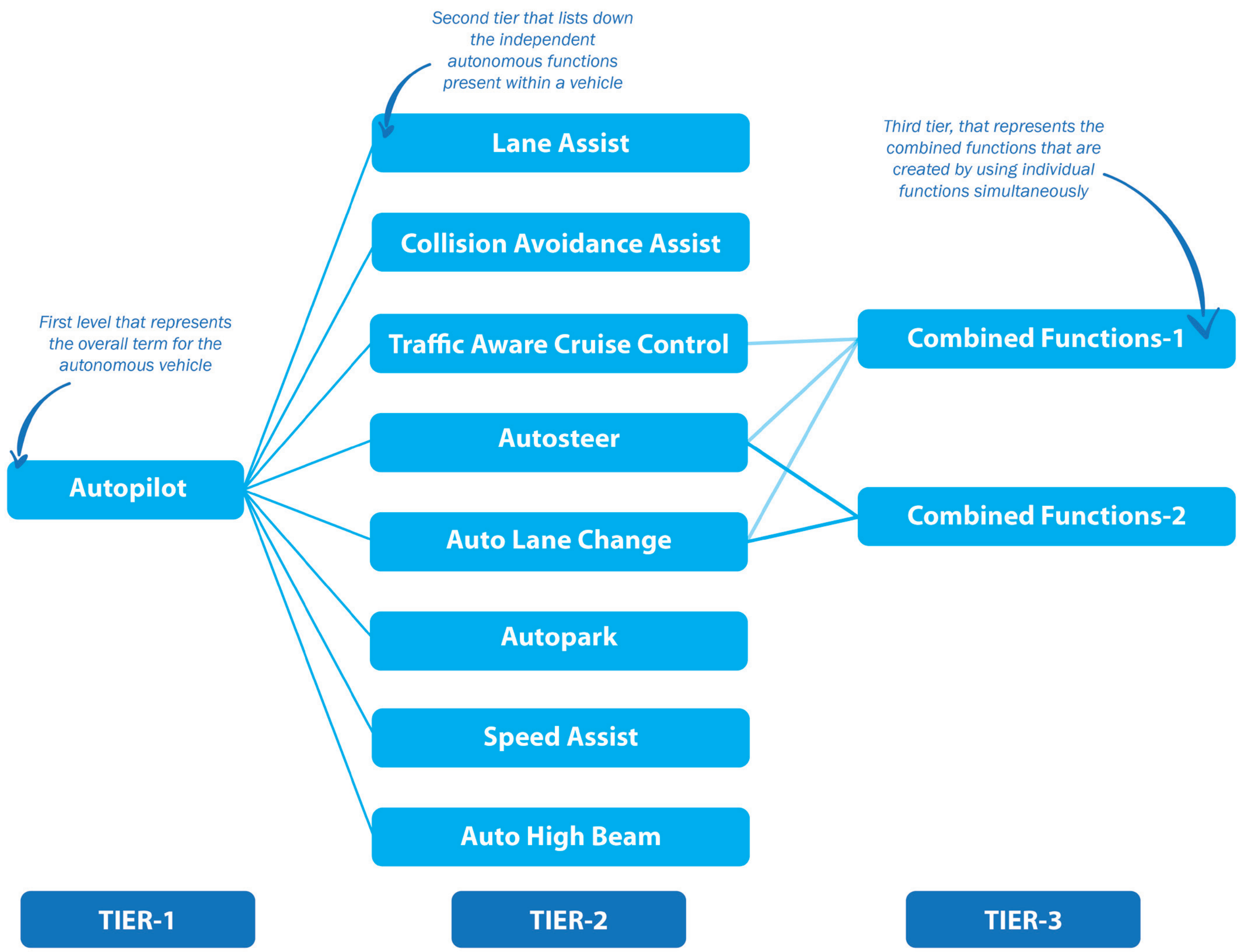


Figure 5.9. The function tree for a Tesla Model S, autonomous features as described in the owners manual

Visualization Canvas

The canvas consists of three parts, the first is the basic description of the autonomous function that we want to visualize and its relation with other autonomous functions. The second part is the sensor position and working of the vehicle. It should be noted that in this version of the canvas the sensor list and position have been added but additional technical information with regard to sensors can also be incorporated. The last part is the breakdown of the autonomous function into three parts:the technology used, the context of use and the user behavior. As seen in the canvas there is also a Light Side and a Dark Side, these represent the proper use and improper use of the automation.

The visualization canvas is filled from top to bottom. First the basic idea of the function is described followed by highlighting its position within the function tree. Once this is completed we move on to detailing the three parts that describe the autonomous function i.e. Technology, Context and User. First the technology questions are answered followed by the context and the last part to be filled is the user. Within the user section there is a second section highlighted in blue that raises the question “What the driver would do?”, this question is added to present a neutral overview of the driver behaviour, which can incorporate desirable user behaviour as well as detrimental behavior.

The first section specifies what function we want to visualize and how it relates to the other functions

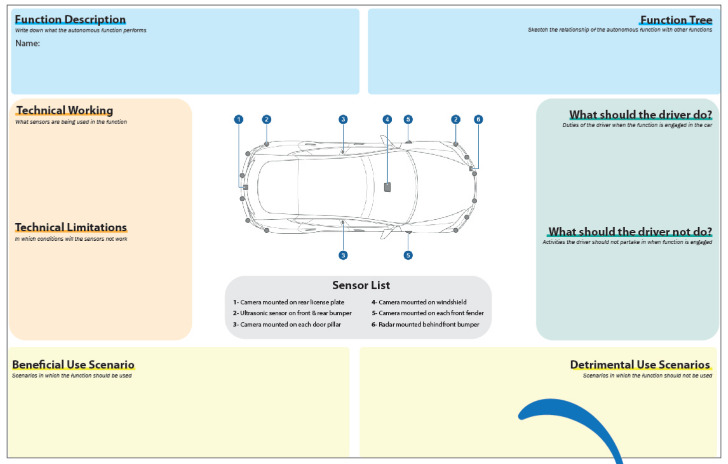
Specifies the sensors present in the vehicle that are used to complete the function

The bottom section goes further into detail about the technical working, user scenarios and user behavior

The section is divided into two parts the light side and the dark side to represent the proper use and improper use

Figure 5.10. Description of the autonomous function visualization canvas

Iteration Process



Iteration-1

The first iteration covered all the major points and during testing the participants were able to successfully fill the canvas. However, there was confusion in the layout of the canvas the the starting point.

Iteration-2

The layout was modified in the second iteration along with providing a more clear distinction between the proper use and improper use of the automation

Iteration-3

The final iteration some small changes were made like adding the function name and the section of what user might do

Using the Autonomous Function Visualization Canvas

The procedure of using the Autonomous Function Visualization Canvas is as follows:

- *Create the Function tree or perceived function tree listing down all the autonomous functions in tier two and the combined functions in tier 3*
- *Select one function from the function tree that you want to visualize*
- *Using the canvas fill in the details of the autonomous function*
- *Repeat step 2 and 3 for every iteration of the autonomous function*

Special Case: In case there is a change in the autonomous function tree such as an addition of a new autonomous function or elimination of one, then for that iteration step 1 will also need to be repeated. For example, if a new autonomous function is added then the tree will need to be modified, following that step 2 and 3 in the process will be repeated.

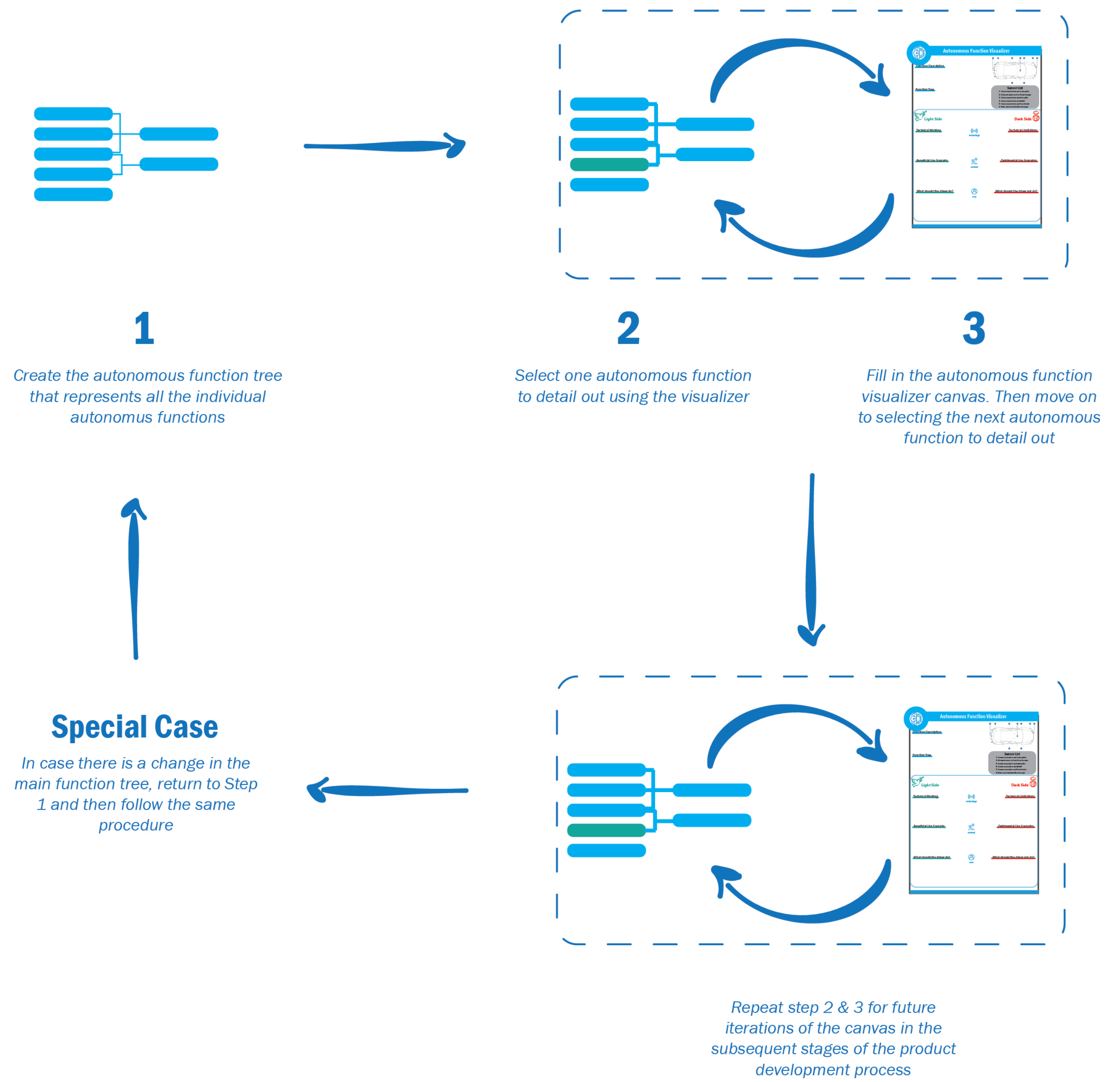


Figure 5.12 A representation of the use of the autonomous function visualization canvas

Position in the Product Development Process

As described in the introduction of the toolkit, the autonomous function visualization canvas is used in different phases of the product development process. In this section we explore the individual phases in which it is used and the reason for selecting those phases:

Concept Development: The first time the autonomous function visualization canvas is used is within the concept development phase in the "Establishing Target Specifications". This is the stage in which the initial specifications for the vehicle prior to the ideation phase. The visualization canvas allows for creating a rough draft of the autonomous system the vehicle would have. The next stage in which the canvas is iterated upon is the "Selecting Product Concepts", this is the stage in which the final concept is decided. The last part of the concept development process in which the canvas is iterated upon is the "Set Final Specifications". This is an important stage as, within this stage all the specifications of the vehicle are specified for the final design and very little to no change occurs in those specifications.

System Level Design & Detail Design: In the concept development phase, the complete design of the vehicle is considered. However, in the system level design and detail design the individual systems work separately and simultaneously to meet the final specifications. In this phase the visualization canvas is further iterated and modified depending on the evolution of the technology.

Testing & Refinement: The last phase of using the visualization canvas is the testing of the complete autonomous system with the other systems. During the testing phase the final version of the autonomous function visualization canvas is created for that particular autonomous vehicle.

The reason for having a number of iterations of the visualization canvas is due to the duration of the product development process for automobiles. Nearly, 60% of automobile manufacturers take around 24 months (Morley, 2019) to develop a vehicle. This is a long duration from conceptualizing a vehicle to reaching the market. The visualization canvas allows for capturing the development of the autonomous system in a systematic manner, mankind it easier to communicate the information at each phase to other stakeholders.

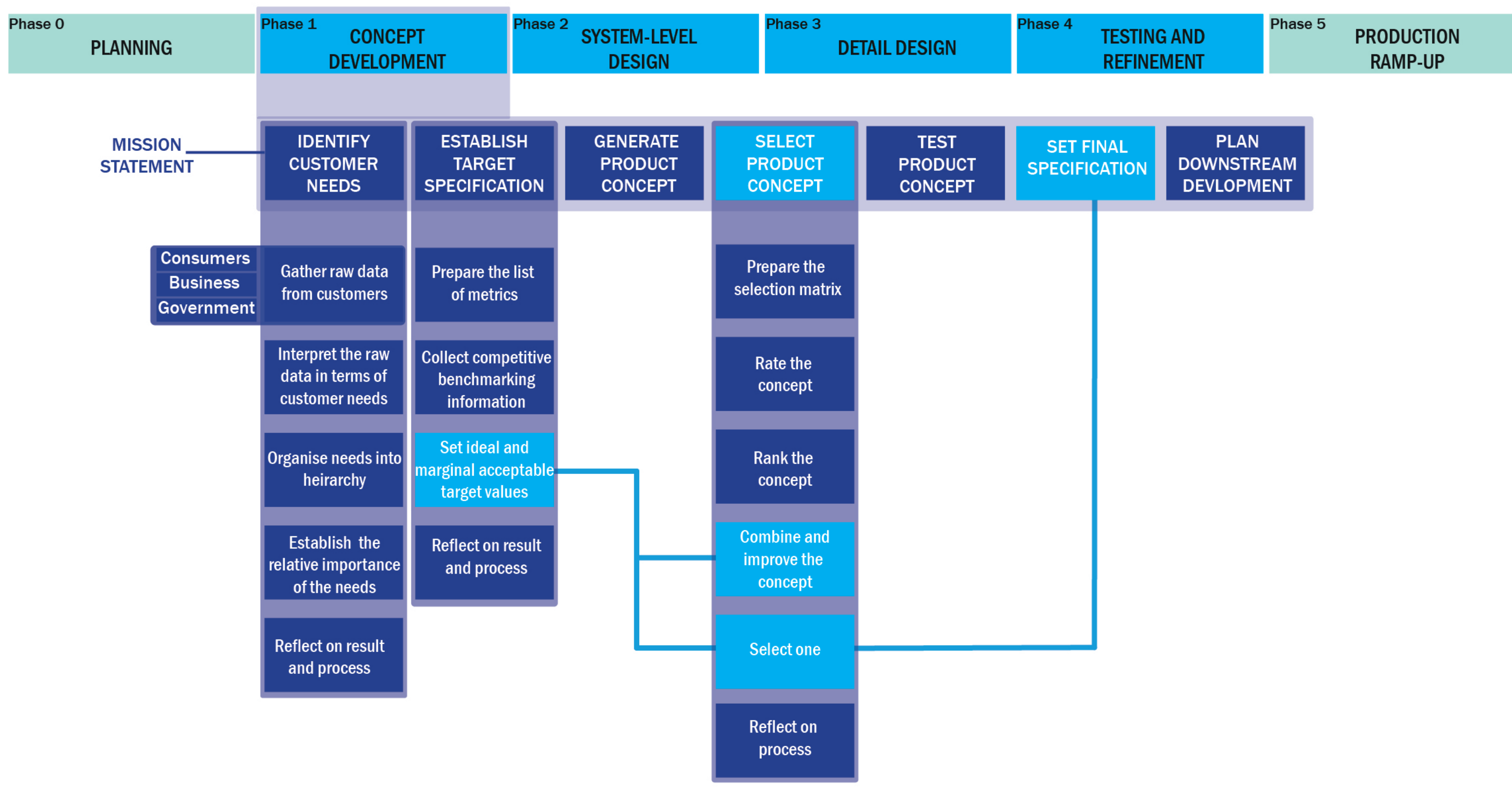


Figure 5.12. The position of the autonomous function visualization canvas within the product development process

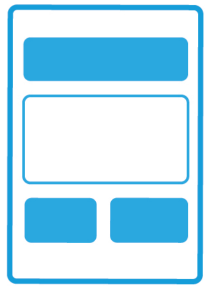
User Decision Matrix

5.4 | User Decision Matrix

The Users Decision Matrix is used to understand the context of use and how it influences the decision people make while driving. It dovetails the autonomous function visualization canvas by allowing the development team to explore the driving task within which the autonomous function would be engaged. While the visualization canvas lists down the proper and improper use based on the context, user behavior and technology. The user decision matrix facilitates the exploration of the influence variation in the context or user behaviour has on the interaction with the autonomous function. The user decision matrix is based on the work by Visser et al.(2014) and Mirning et al. (2016). The matrix is created by considering the decision making process of a driver along with the factors that are present in the context of use such as other road users. The insights captured are focused towards the development of HMI and eHMI systems for autonomous vehicles. The user decision matrix like to Autonomous function visualization canvas has two parts: the first is the scenario sheet and the second is the user decision matrix.

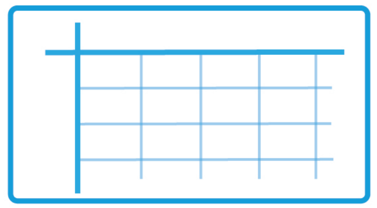
The scenario sheet is designed to understand the context of the driving task that we want to analyse. It is based on the Situational Trust factors as proposed by Hoff & Bashir (2015) and allows the whole team to create a common understanding of the driving task being analysed and the context of use, by explicitly highlighting and visualizing them in the Scenario Sheet.

The User Decision Matrix is the subsequent step after completing the Scenario Sheet. It is a 4X3 matrix, with the stages for making a decision (Perceive, Understand, Predict/ Perform and Adapt), on the X axis and the situational factors (Other road users, context of use and my vehicle) along the Y axis. The Decision matrix should be only be used to analyse a single driving task at a time example, taking a turn or overtaking a car at a time and multiple driving tasks must not be combined into one scenario as it dilute the outcomes. The user decision matrix can be used in association with the autonomous function visualization canvas or as a separate tool altogether. The various scenarios of use have been illustrated in the latter sections.



Scenario Sheet

The scenarios sheet allows for visualizing scenarios. It consists information about the scenarios such as the action the vehicle needs to perform along with context factors and driver insight. The sheet makes it easier to envision the action under analysis in the user decision matrix.



User Decision Matrix

The decision matrix consists is created by considering the steps of decision making along the x axis and the context factors along the y axis. The Matrix allows for breaking down each action taken into its decision steps that are in line with the various context factors (road users, road condition etc.)

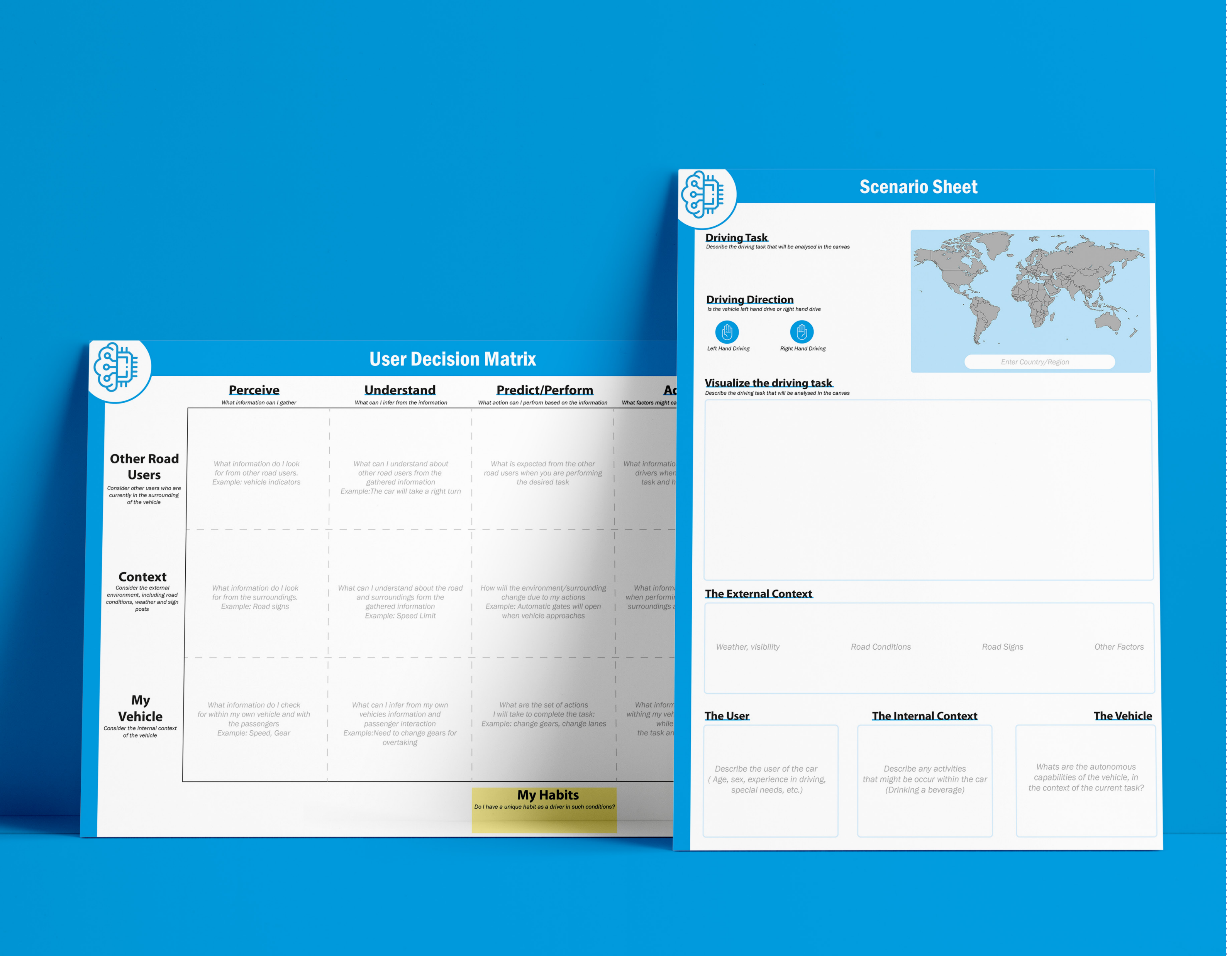


Figure 5.13. Mockup of the Scenaro Sheet and User Decision Matrix

Scenario Sheet

The Scenario Sheet is designed to understand the context of use and is based on the Situational Trust aspects of internal situational trust and external situational trust.The scenario sheet is divided into four major parts. The first part is the description of the driving task that is intended to be analysed, this further includes the region in which the vehicle is believed to be used and the driving direction (left hand or right hand). The second part is the visualization of the scenario, this can be done by models, sketches or pictures, any medium that makes the scenario clear to the team. However care must be taken to not make the scenario too generic as it leads to loss of valuable insight. The third part is the description of the external context, which includes the road conditions, weather, visibility etc. The last part is the internal context which consists of the vehicle we want to analyse, the characteristics of the driver and the internal context of the vehicle (passengers, auxiliary activities being performed, etc.). While initially the internal context of the vehicle might not seem as an important factor influencing the driving task. They exert a major influence on our decision making. The challenge is to identify these internal context factors. Collectively the four make the scenario sheet which provides a holistic and detailed overview of the driving task that is to be analysed.

While the scenario sheet provides a systematic representation of the required driving task. Its main importance is on building a common picture within the team members when using the decision matrix. A vague representation of the context invites more assumptions. Having assumptions is not a bad thing but if every member of the team has different assumptions while filling the user decision matrix, it will not bring any meaningful insight. Thus, it is important to detail the scenarios sheet in a systematic manner to be create value when using the user decision matrix.

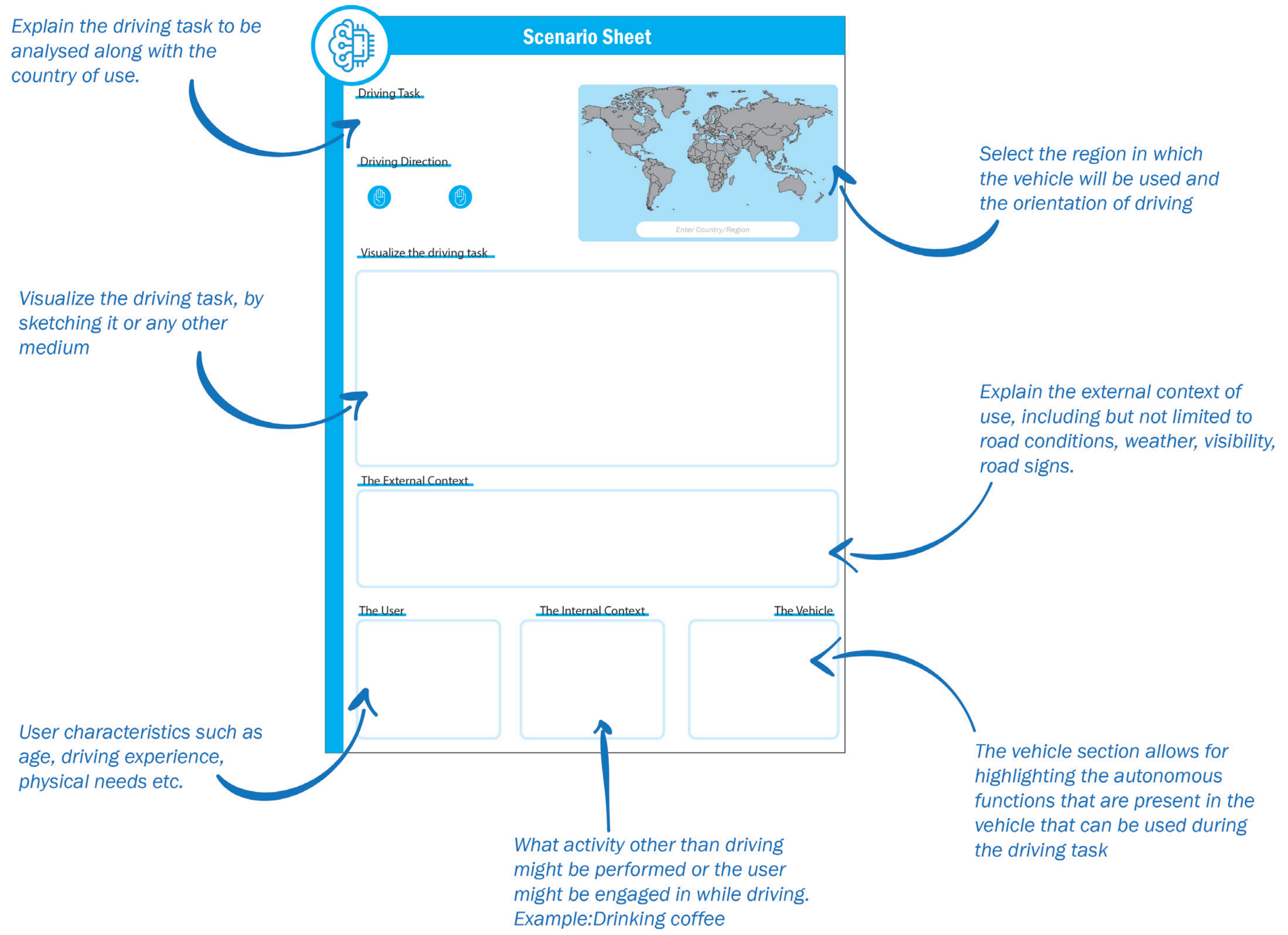


Figure 5.14. Breakdown of the scenario sheet

User Decision Matrix

The user decision matrix as stated earlier is a 4X3 matrix. The matrix is filled from top to bottom and left to right, there is however one exception when filling in the Predict/Perform column which is filled from bottom to top. The first column in the matrix is that of Perceive which consist of the information cues we gather from the context and other road users. The second column is Understand, at this stage we make sense of the information gathered in the Perceive stage. This is followed by the Perform/Predict stage in which represents the steps taken to accomplish the task. The last part is the Adapt stage, which represents how the user would adapt to a change in the context when performing the driving task.

Along the Y axis we have the Other Road Users in the first row, the second row consists of the Context. The last row is for the user and his own vehicle titled My Vehicle.

Once the user decision matrix has been completely filled, the can then be used to understand what information changes occur when there are changes in the context (external and internal). These are then further researched to see if these changes influence the users decision making or trust in automation. In essence the user decision matrix provides a systematic method to look at the effect the change in context has on driving decisions.

Using the User Decision Matrix

Three situations have been identified within which the user decision matrix should be used. This however does not mean that there do not exist other scenarios in which the user decision matrix can be used. The three scenarios are as follows:

- **Influence of specific context:** During detailing the design of HMI systems or eHMI systems, the user decision matrix is used to vary the context both externally and internally to understand how that would impact the users decision making process and what information is relevant at that time. This is achieved by selecting a particular autonomous function and a specific context factor that needs to be studied. First one scenario sheet and user decision matrix is filled. Following this the context factor is varied and the step is repeated. Once there are two decision matrices. The variation in the context factors can be explored.
- **Assumption Verification:** During testing of the autonomous vehicle the User Decision Matrix is used for comparing the assumptions the design/ development teams made and how it varies in an actual scenario or a simulation. This is achieved by comparing simulation data for a specific driving task to the assumptions made by the development team as filled in the user decision matrix.
- **Designing for Edge Cases:** An autonomous vehicle facing an edge case is one of the prominent concerns for designers and developers as it is difficult to predict what the outcome of that interaction would be. The phrase edge case refers to scenarios that are novel to the autonomous vehicle such that they are unable to accurately determine the course of action. As seen from Figure 23, an edge case can be analysed in retrospect, by breaking down the users decision making and analysing the sequence of actions.

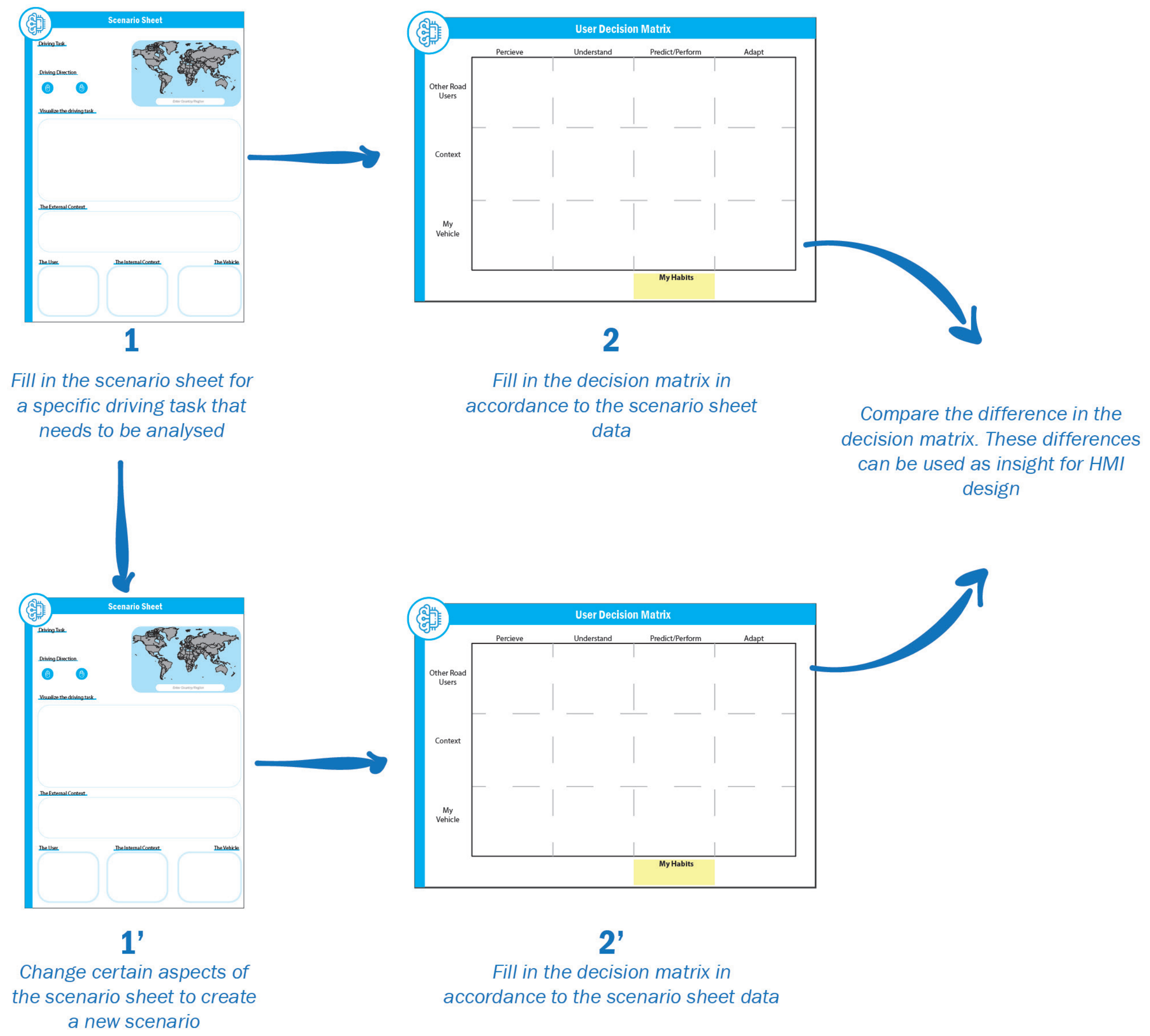


Figure 5.15. Process of using the user decision matrix in case of studying influence of specific context factor

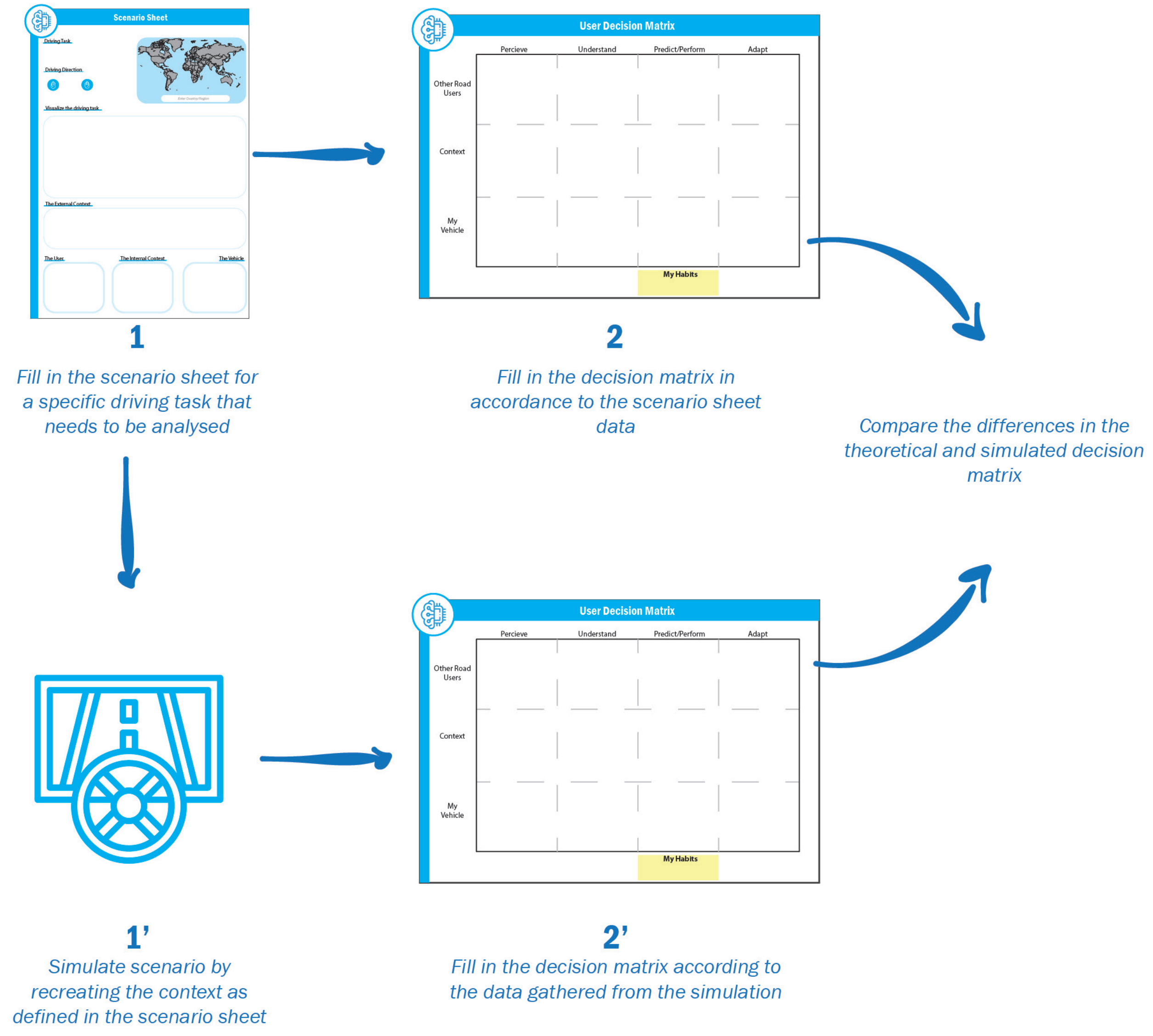


Figure 5.16. Process of using the user decision matrix in case of studying assumption verification

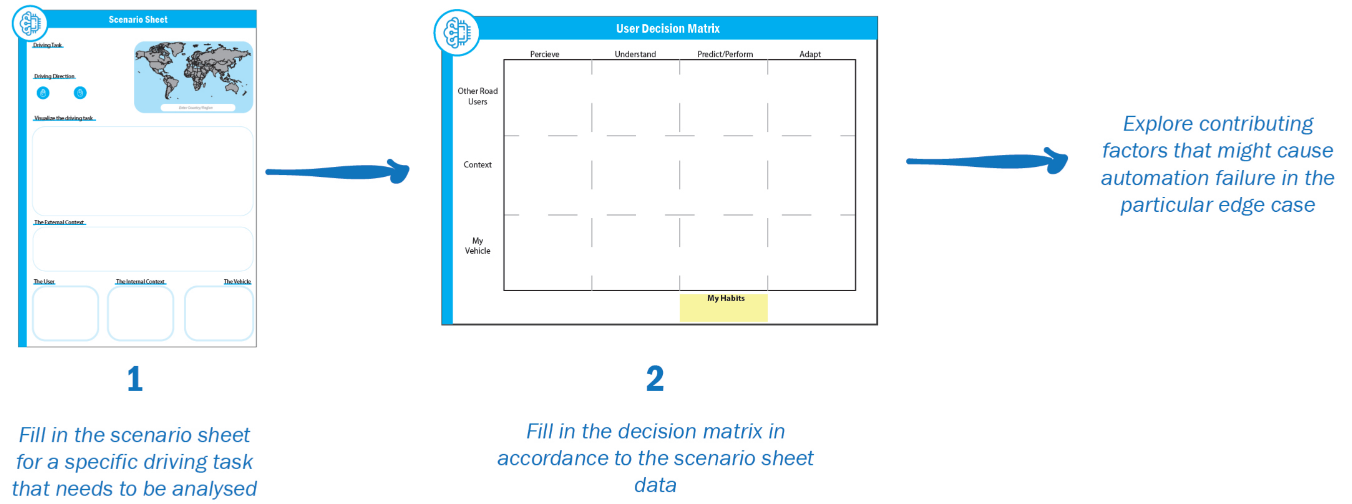
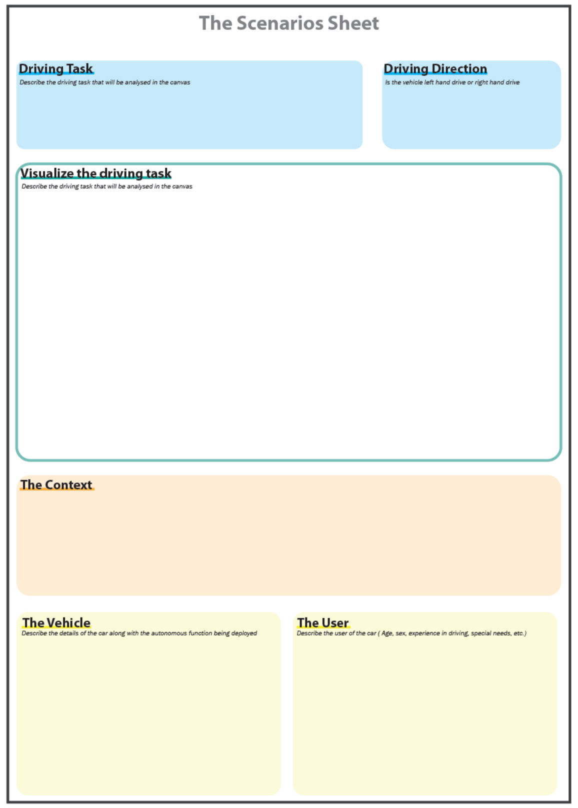


Figure 5.17. Process of using the user decision matrix in case of studying edge cases

Iteration Process

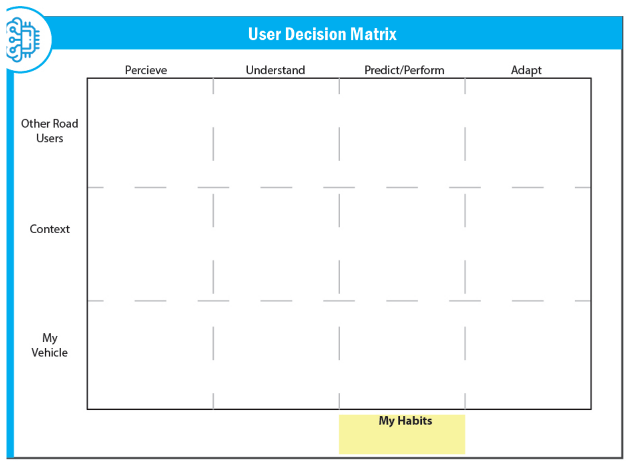


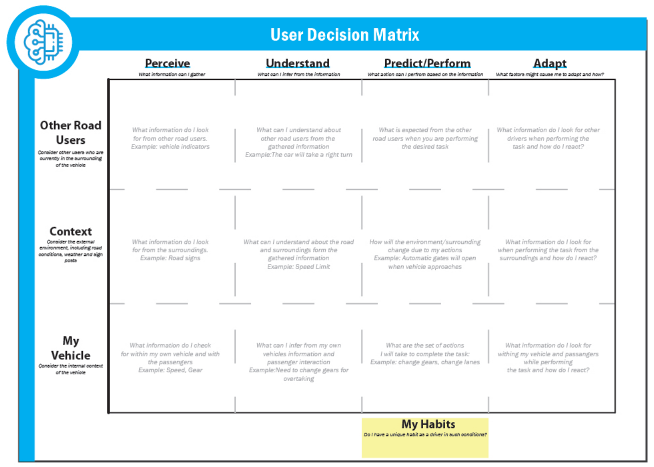
Iteration-1

The first iteration covered the external context and very limited internal context of the vehicle

Iteration-2

The second iteration covered the internal factors as well as the external factors. Providing a more complete overviw of the driving task





Iteration-1

The user decision matrix initially only consisted of the matrix and the direction and details of using the matrix were manually explained

Iteration-2

The second iteration examples were added to the individual parts of the matrix to make it more intuitive. This was the only change made

Positioning in the Product Development Process

The User Decision Matrix is to be used in the System Level Design, Detail Design, Testing and Refinement phases of the product development process. This means that the user decision matrix is introduced much later within the product development process as compared to the other parts of the toolkit. This is because it is beneficial to use the user decision matrix after the final specifications of the vehicle have been decided upon.

- **System Level Design & Detail Design:** Within this phase the user decision matrix is used to explore the influence of context and assumption verification (described in the previous section).
- **Testing and Refinement:** Within the testing phase, assumption verification is the prominent scenario for using the decision matrix. In addition the use to analyse Edge Cases is also an option. However, for this the testing needs to create an edge case or encounter an edge case so that it can be analysed.

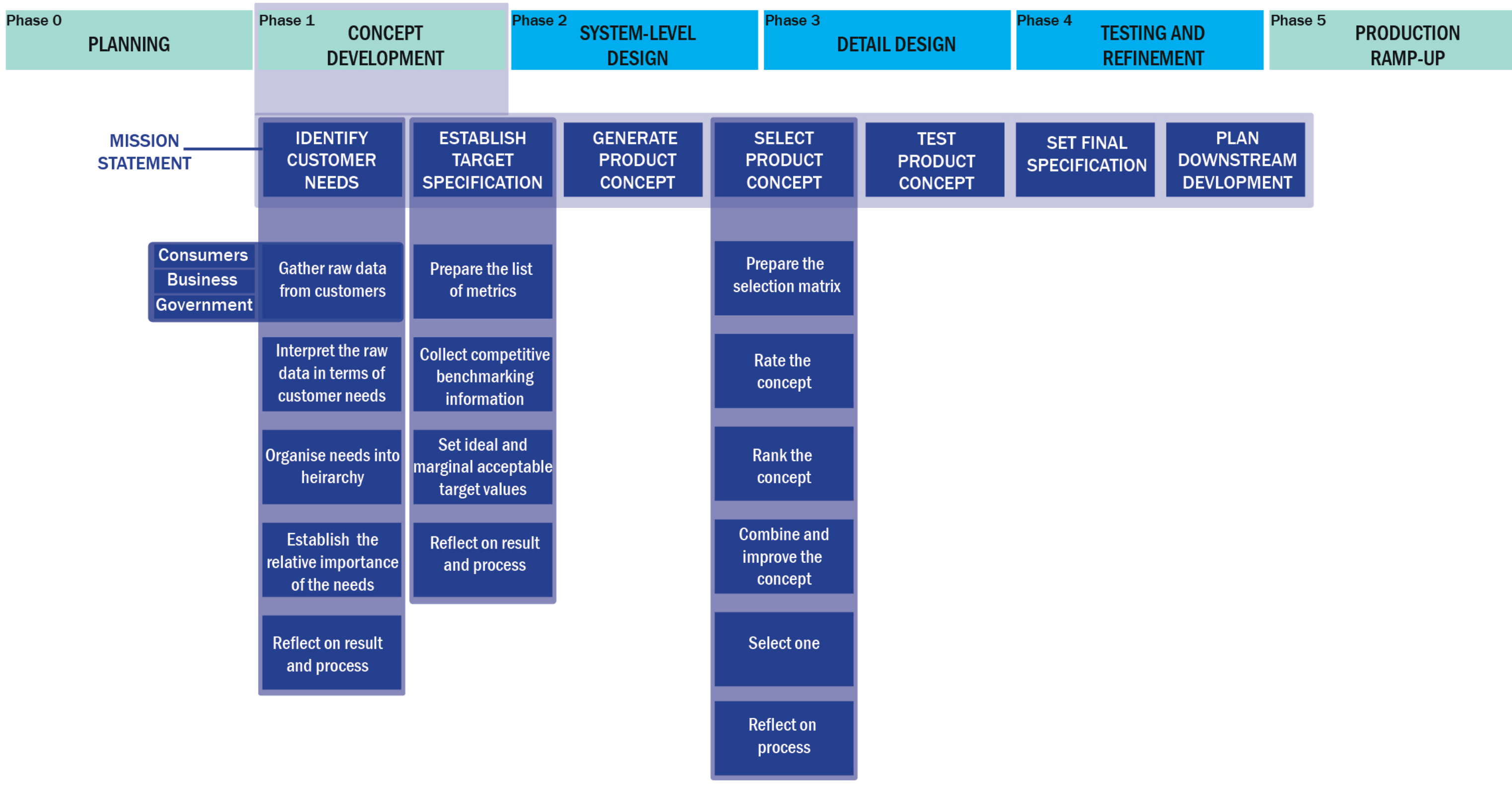


Figure 5.18. Position of the user decision matrix along the product development process

Trust Enhancing Communication (TEC)

5.5 | Trust Enhancing Communication

The last part of the design toolkit is the Trust Enhancing Communication. As discussed in the formulation of the new design question, communication plays a major role in calibration of trust. The sensitizing session aids in creating an understanding of the basics of calibration of trust. The autonomous function visualization canvas and the user decision matrix allow for operationalizing of calibrated trust. TEC is the last part of the set allowing for effective communication of information about the autonomous system to stakeholders both internal and external.

Trust Enhancing Communication (TEC) is based on the concept of Trust-Enhancing Technologies (TET) (Withworth & De Moore, 2009). TET defines a set of properties that support and improve the assessment of confidence between people and digital media. Trust Enhancing Communication takes certain parts form TET and modifies it to meet the context of communication for autonomous vehicles.

TEC consists of 3 principles which act as the foundation of this method of communication. The three principles are interrelated and the communication must obey all the three principles:

- **Amplification, not alteration:** TEC amplifies evidence of trust and calibrating trust evidence. This is achieved by isolating them from other information or noise, visualizing them for better communication or aggregating them across large data collections (supporting them by statistics)

- **Transparency, not blind trust:** TEC does not support the creation of new evidence and does not interfere with the interaction between the user and information. All evidence of trust is available to the user and within their reach to explore

- **Better assessment, not greater confidence:** TEC supports the right use of communication of autonomous functions. Making the spreading of improper information challenging and creating barriers for checking.

The TEC principles do provide an overview of what is expected from communication. But to achieve these principles there exist certain requirements that can shape TEC. These requirements are perspectives but not definitive and can be implemented in several different technical manners (Cofta, 2009). These requirements act like a framework to discuss and compare the different methods/modes of communication. It however, should be pointed out that the requirements must be considered jointly and the final result must satisfy all of them.

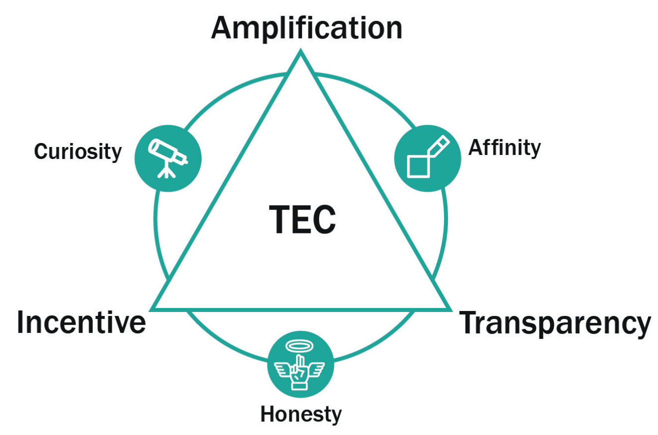


Figure 5.19. Overview of Trust Enhancing Communication

- **Affinity:** The requirement addresses the need to disclose the cognitive and intentional capabilities of the autonomous vehicle. Affinity is not concerned with what interpretation is created by the sharing of the information, but only the fact that the information be disclosed to the stakeholders.

- **Honesty:** It is the requirement to tell the truth and the fact that communication channels do not distort, damage or alter the information.

- **Curiosity:** It is the ability to interact and confirm the information by stakeholders. This means that the information is communicated in a manner that allows for a certain level of interaction. This can be accomplished by prototypes, simulations or other modes of communication.

TEC can be used as a source of inspiration or a point of scrutiny. During the development of the communication material it can be used as a source of inspiration to think about novel methods of communicating information. In the latter stages it can act like a check to see that the communication material meets the requirements. TEC must not be seen as individual principles or requirements but as a collective whole.

Using TEC

The procedure of using TEC starts by satisfying the three requirements starting with affinity and ending with curiosity. Subsequently, we move to the three principles. For the principles we are free to choose which one to start with as they all are interrelated and need to be satisfied. The steps for implementing TEC are as follows:

1. Choose which stakeholder is the recipient of the information
2. Select the information that needs to be communicated e.g. details of an autonomous function
3. Start by checking the information for the satisfying the requirements

   Affinity: Collect all the relevant information about the autonomous system, no information should be withheld as it is up to the recipient to create their understanding of the autonomous system.

   Honesty: Arrange the information in a manner that does not distortion or alter the information.

   Curiosity: List down the various modes of communication (boundary objects), the list must not only consist of conventional communication methods but also methods that allow for interaction with the information such as prototypes, provotypes or sensitizing sessions (NOTE: This is the one part of the TEC that require thinking out of the box, it is easy to satisfy the first two requirements)

4. Having completed the requirements we move on to the principles, as we are free to choose any principle to start with. The procedures are provided for all three:

   Amplification: Arrange the information in a manner that allows for prioritising information that focuses on calibration of trust/ creating the appropriate mental model.

   Transparency: Make sure to provide all the relevant information to the stakeholder. This goes hand in hand with the first principle of amplification. In amplification we organise the information, whereas transparency acts like the check to not miss out on any relevant information.

   Incentive: This principle can take two forms. The first is creation of barriers to prevent misinformation or incorrect interpretations. The second is incentivisation of stakeholders to interact with the information. Depending upon the stakeholder either of the two forms of incentive can be user or both can be considered simultaneously.

To illustrate the use of the TEC, we consider a situation in which the marketing department requires information with regard to an autonomous system for a new vehicle that is about to be launched. We explore through this example the use of TEC when interacting with internal stakeholders.

1. *We have our recipient of the information in the from of the marketing team*
2. *The information that needs to be communicated is the capabilities of a specific autonomous function*
3. *Look at the three requirements, as we are communicating an autonomous function we can use the autonomous function visualization canvas*

   ***Affinity**: The autonomous visualization canvas contains all the relevant information about the autonomous function*

   ***Honesty**: Check the information within the autonomous function visualization canvas to see that it is clear and will be understood by the recipient*

   ***Curiosity**: List down possible modes of communication. In this case we can use the visualization canvas itself, a simulation of the autonomous function with a written document or a real time video recording of the working of the autonomous function along with a brief description. In this case we choose the visualization canvas and a real time video recording of the working autonomous function as our mode of communication.*

4. *Moving to the principles of TEC*

   ***Amplification**: According to the needs of the marketing department, the relevant information within the visualization canvas are highlighted to make it prominent. Additional relevant information is added so that the recipient can understand the complete use of the autonomous function, which is also helped by the video.*

   ***Transparency**: The information is rechecked to see that all information is provided in a manner that is understandable to the recipient and not information is modified or removed.*

   ***Curiosity**: The last part of the principle would be to create a barrier to prevent misinformation or misinterpretation of information taking place. To overcome this a follow up meeting can be scheduled with the marketing department. This allows for exploring that the marketing department has interpreted the information in the correct manner and any misinterpretations can be clarified.*

Figure 5.20. Mockup of the trust enhancing communication in the form of a poster to be distributed

This example highlights that to apply TEC there is no need to create additional information or boundary objects, but just ordering them in the correct order can provide a great dividend. It also highlights the simplicity of using TEC when used in conjecture with the other parts of the toolkit.

The TEC is used during and after the concept development to communicate system information about the autonomous functions. As there is no limitation to a specific location of using TEC we will not be focusing on the positioning of TEC specifically within the product development process.

Part 3 I Chapter 6

Testing & Validation

This chapter details out the testing and validation process employed within the project. The first section of the chapter is focused towards the individual parts of the toolkit and the testing/validation of these individual parts. The second section is the validation of the complete concept which was performed by interviewing experts.

6.1 I Validation of Individual Parts

The first part of the test and validation provides an overview of the procedure and insights that were recorded for the individual tests. Three parts of the toolkit were tested: The sensitizing session, autonomous function visualization canvas and the user decision matrix. The TEC was validated during the concept validation which has been discussed in the next part of the chapter. We begin first with the sensitizing session, followed by autonomous function visualization canvas and the last part is the user decision matrix. The test setup and the requirements for each of the three parts have been explained individually within their respective sections.

Sensitizing Session

Test & Validation Setup

The testing and validation of the sensitization session was performed by conducting two sessions. Each session was attended by two participants. The participants were provided a sensitizing package a week before the session. This was in a digital format that made it easier to fill in and share during the COVID-19 circumstances. The average time required to fill the sensitizing package was estimated to be 30 minutes. The participants were required to send the sensitizing package back to the researcher a day before the actual sensitizing session. The participants were then invited to the online sensitizing session, which lasted for about 2 hours (the duration was less than the actual time proposed because of less number of participants).

After the completion of the sensitizing session, an semi structured interview with regard to the overall session was conducted. Participants were questioned about both the sensitizing booklet as well as the session. The insights gathered during the testing of the design were incorporated in the next iteration of the design and tested again.

A second round of validation was performed by analysing the responses of the participant had provided during the session, in the form of Post-it notes. The responses of three canvases selected to analyse the “Trust Goal” reflected their understanding of calibration of trust. The three canvases selected were: Trust In Automation, The Process of Trusting and Trust Goal. The reason for selecting these three canvases is because they focus on trust between human and automation and are relevant to the aim of the sensitizing session.

Requirements for Sensitizing Session

- *The session must provide an overview of the factors that influence trust*
- *Create an understanding of calibrated trust*
- *Convey the understanding of trust to other stakeholders by means of the trust goal*

Result

The participants agreed that the sensitizing session forced them to think about trust and also try to ground an abstract idea like trust into more tangible concepts. They also agreed that their understanding of trust has changed as highlighted by the quote, “*This whole session has oriented me to think in a certain way, and its brought some kind of awareness that I did not have before this session. The whole idea of being gullible that we discussed before, with new knowledge I think I will be less gullible*”. They appreciated the exploration of factors influencing trust and enjoyed the contradicting or conflicting ideologies presented by the fellow participants “*I think you structured it very well to match the chain of thoughts we were having during the session*”. All participants agreed that they now understood the meaning of calibrated trust or appropriate trust, but interestingly also highlighted that achieving it is a tough challenge for designers as it is an ideal scenario. The creation of the trust goal was appreciated as a conclusion to the sensitizing session “*I think it is a good goal to have, and it has a lot of things that we discussed, I know with time it will improve, but the goal was pretty close*”.

The creation of a trust manifestation (Boundary Object) in addition to the trust goal did not go as planned as the participants in both tests misunderstood the aim of the trust manifestation and felt that it was not beneficial or thought provoking. For this reason the trust manifestation was removed from the final design and the sensitizing session was ended with a trust goal. Overall, the participants did not find the session boring and enjoyed the discussion and activities. Thus, suggesting that the requirements set for the concept were validated.

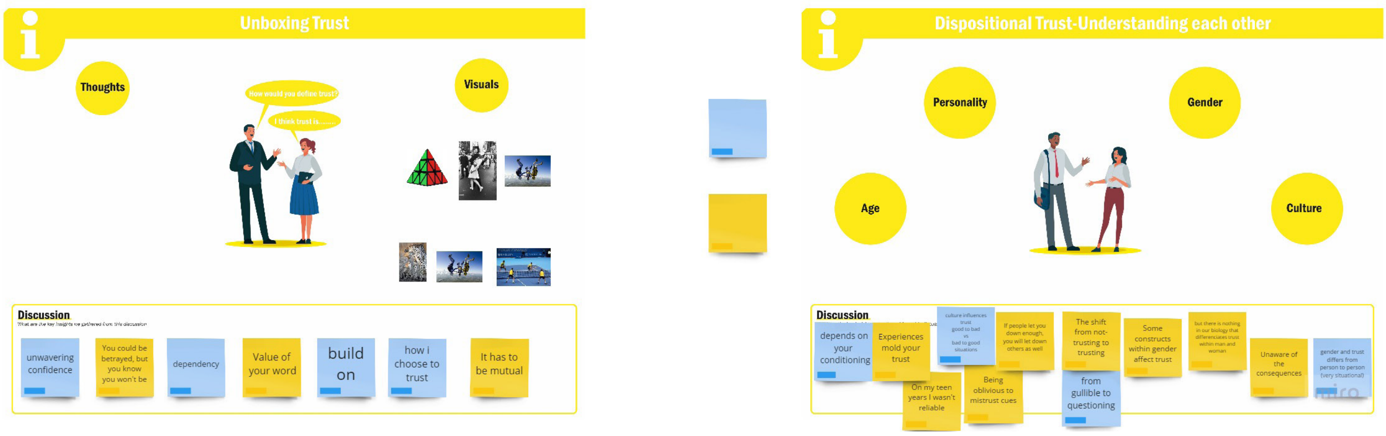


Figure 6.1. A snapshot of the filled canvas of the sensitizing session

Insights for the Sensitizing Package
The sensitizing package was like and appreciated by the participants the most. They felt that it provided them time to think about trust and also record their thoughts. In addition the use of images and words to stimulate their answers was highlighted as one of the benefits of sensitizing booklets in describing an abstract idea as seen in the comment "*To think about an abstract concept it was helpful to have pictures, and some of the words really made me think about trust*" . The workload was not considered a hurdle in their day to day activity and the digital version of the booklet was easy to fill in. None of the questions were confusing or unclear to the participants. However, the frequency of filling in the booklet was seen to vary, while some filled in the booklet in multiple sessions others did it in one sitting. This made gauging the actual time they spent on the booklet a little challenging.

> *"To think about an abstract concept it was helpful to have pictures, and some of the words really made me think about trust"*
>
> -Participant 2 (Session-2)

Insights from the Sensitizing Session
The overall sensitizing session was seen as helpful by the participants. However, they did highlight certain canvases that were not as intuitive and easy to use. The "Trust Process" canvas was considered as one of the challenging canvases because they were not familiar with the process of trust formation. They suggested that adding examples in the canvas or during the introduction of the canvas would help in understanding the canvas better. In addition the Trust Goal canvas was a little unclear to the participants in the beginning "*Yeah I dont think its about the instructions, the last canvas was pretty abstract and so it was difficult to find my opinions*". The participants like the "Trust Misconception" canvas as it challenged some of their presumptions as seen by the quote "*Definitely, especially the part where you said how trust should go, and the part where you were shattering the misconceptions.... It was really nice to get back something from the process as well*". They also appreciated the "Calibrated Trust" canvas in which they were required to physically plot their trust levels and then discuss what should be the ideal trust level.

> *"Definitely, especially the part where you said how trust should go, and the part where you were shattering the misconceptions.... It was really nice to get back something from the process as well"*
>
> -Participant 2 (Session-1)

The second round of validation which was based on analysing the responses of the participants was conducted after analysing the interviews the participants had given at the end of the sensitizing session. The subsequent paragraphs describe the results of the second round of validation.

The Trust in Automation canvas acts like the benchmark for understanding the initial idea of calibrated trust the participants have Figure 30. The orange dots represent the participants initial response, while the blue dot represent the ideal response as perceived by the participant. As seen in both the sessions not all participants' responses collaborated with the calibrated trust level (represented by the diagonal black line). The reasons for the deviation of trust from the calibration line is further explored in the "Formation of Trust" Canvas. The canvas breaks down the reasons along the trust formation process(Performance, Process and Purpose) as seen in Figure 31.

After completing the exploration of the reasons for not achieving calibration of trust, within the "Trusting Process" Canvas, we move towards the "Trust Goal" canvas. In the final canvas the participants were instructed to highlight key factors that influenced their trust/distrust and form a trust goal moving forward. In session one the trust goal created for Zoom was "***Putting the user at the center of transparency through interactions***". This goal was based on the insights that Zoom was not transparent (process) about their security. In addition there were an abundance of features which made it difficult for participants to use and understand the overall purpose and scope of Zoom (purpose). The trust goal formed at the end represents both these shortcomings that negatively influences trust. This is also corroborated by the insights from the interview where the participant believed the goal was able to capture the reasons for lack of trust in the technology.

Similarly in the second session, the technology chosen was Skype and the trust goal created was "***Skype for anyone in anyway***". The trust goal represented the shortcoming of the performance aspect of Skype. It was pointed out during the session that Skype would crash unexpectedly during a meeting. However, on the contrary the simplicity of the interface was praised along with the purpose. The trust goal in that regard represents the improvement needed to be made in the performance of Skype, which is why the phrase "*anyone in anyway*" was selected as it represents irrespective of the conditions Skype should be able to connect people.

The analysis of the responses of the participants during the session along with the insights of the interviews post the session show that the participants are sensitized towards trust in automation and are able to identify the reasons for mistrust/distrust with regard to the specific automation/ technology. The formation of the trust goal correlates with addressing the reasons for mistrust/distrust, which acts like the confirmation that the participants have achieved a certain level of understanding with regard to the concept of calibrated trust.

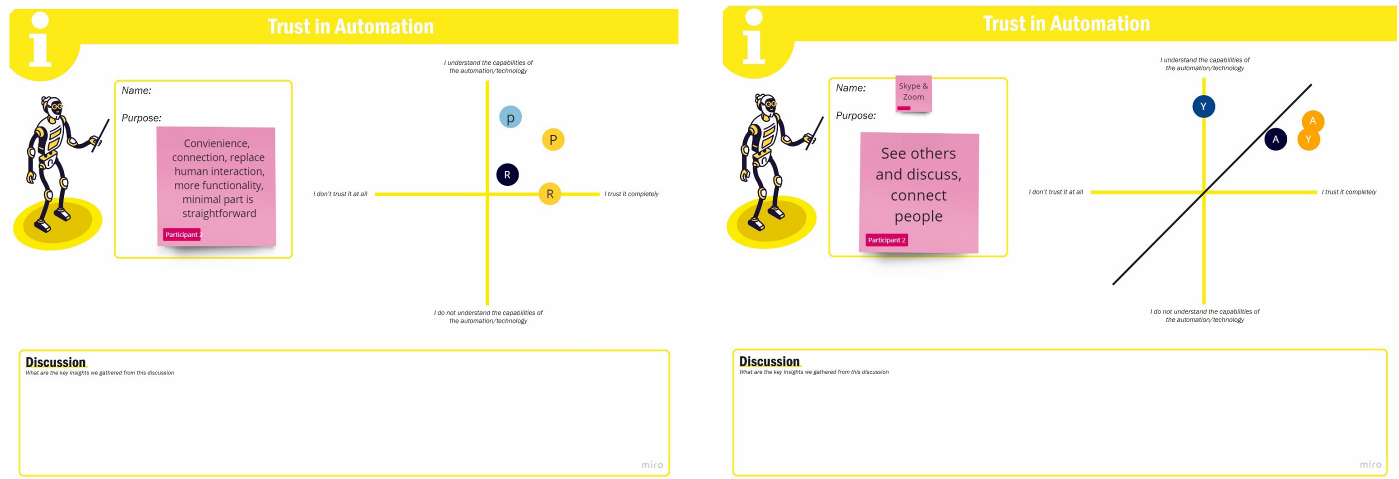


Figure 6.2. The orange dots represent the participants perceived level of trust based on their understanding of the capabilities of the technology/automation. The blue dots represent the participants' guess of what should be the ideal level of trust.

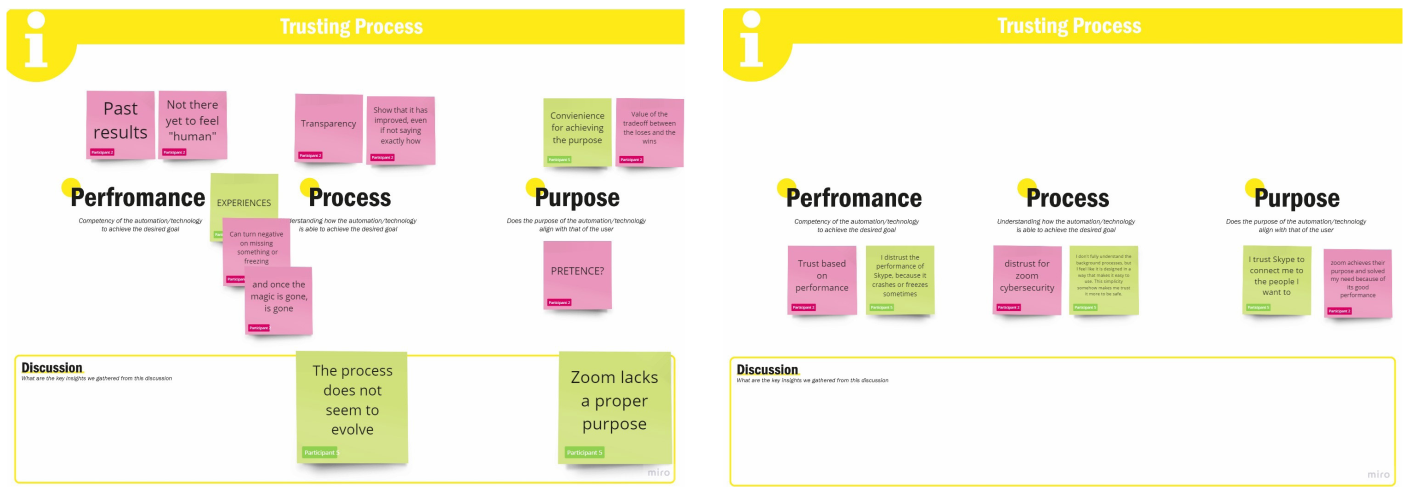


Figure 6.3. Exploring the reasons for lack of calibrated trust based on the trusting process canvas.

Autonomous Function Visualization Canvas

Test & Validation Setup

Three sessions were conducted to analyse the autonomous function visualizer. The sessions were conducted online using Zoom and Miro. For each session the participants were provided a particular function of the Tesla Model S, e.g. adaptive cruise control. The participants were required to read the details about the function prior to the main session. Two separate functions were chosen for this testing, the first was Traffic Aware Cruise Control and the second was Lane Assistance. The participants were required to read through the autonomous functions document provided and watch a YouTube video on the working of the autonomous function. This was done to provide more context and understanding of the function to the participants.

During the session the participants were asked to fill in the canvas using the description of the autonomous function provided to them. The participants were encouraged to think aloud to understand any difficulty they face while filling the canvas or any missing information. Following the completion of the canvas the participants were invited to a semi-structured interview to assess the design intervention compliance with the specified requirements. Participants from different fields and expertise were selected for the three sessions that were conducted. One session was for engineers working/studying in the field of mobility and vehicle design. The second session included designers and the last session was with engineers but without specialization in mobility or autonomous vehicles. This was done to test the ease of using the canvas and its intuitive nature.

The validation of the design intervention followed a similar setup as in the case of the sensitizing session. The first validation is from the semi-structured interviews with the participants. The second is by analysing the data the participants filled in the canvas. This was done by comparing the responses of the participants with the desired information, as filled by the researcher. This allows for studying the quality of the responses and how their responses in the interview corroborate with their filling of the canvas. It also allows for exploring additional shortcomings within the design that might not have been discussed during the interviews.

Requirements for Sensitizing Session

- *Human centered/user centered approach to defining an autonomous function*
- *Understanding of autonomous function and the context of use*
- *Simple and intuitive to use*
- *Represents all the information required to define the working of an autonomous function*

Results

The results of the test and validation were favourable, the participants were able to complete the visualization canvas with the information provided. They acknowledged the structured design of the layout allowed them to enter the information easily and at the same time understand the functioning of the system *"The canvas was well structured and well correlated, the canvas provided a systemic overview that a layman would also be able to understand.....we were able to translate the data in a crisp and concise way onto the canvas"*. During the first testing there were certain challenges that were observed but the participants did not highlight them in the post test interview. This related to the order of filling the information. Participants were confused in which order to fill the information and required some initial guidance with respect to this. These issues were addressed in the second and third iteration by changing the layout and using color to denote the information. This was seen to aid the participants in the second and third test in filling the canvas.

> *"The canvas was well structured and well correlated, the canvas provided a systemic overview that a layman would also be able to understand.....we were able to translate the data in a crisp and concise way onto the canvas"*
>
> -Participant 2 (Session-1)

There were two suggestions that were pointed by participants that have not been added into the current design. The first was the inclusion of more technical knowledge of the sensors into the canvas or creating an additional document *"I think the data was all present, I do not know if this should be in the document, but I would like to see a electronic diagram about how the sensors work would be helpful...but I am not sure the additional data might create a data overload"*. While this is a relevant point and was also later pointed out in the validation interview with experts, the reason for not adding more technical information is to keep the focus on the user and the context of use. Also, there exists the dilemma of how much information about the sensors and technology is enough information. This question could not be answered with the time frame of the project but is a valid extension of the canvas. The second suggestion that has not been considered is the use of metaphors to describe the autonomous function. This was an interesting idea and would definitely make it easier to understand the autonomous function. However, with a single autonomous vehicle containing over 6-8 different autonomous functions, finding a metaphor for each one of them would be challenging. In addition the metaphor should be understood universally and not regionally which would again raise some challenges. For these reasons this suggestion was also not included.

> *"I think the data was all present, I do not know if this should be in the document, but I would like to see a electronic diagram about how the sensors work would be helpful... but I am not sure the additional data might create a data overload"*
>
> -Participant 1 (Session-1)

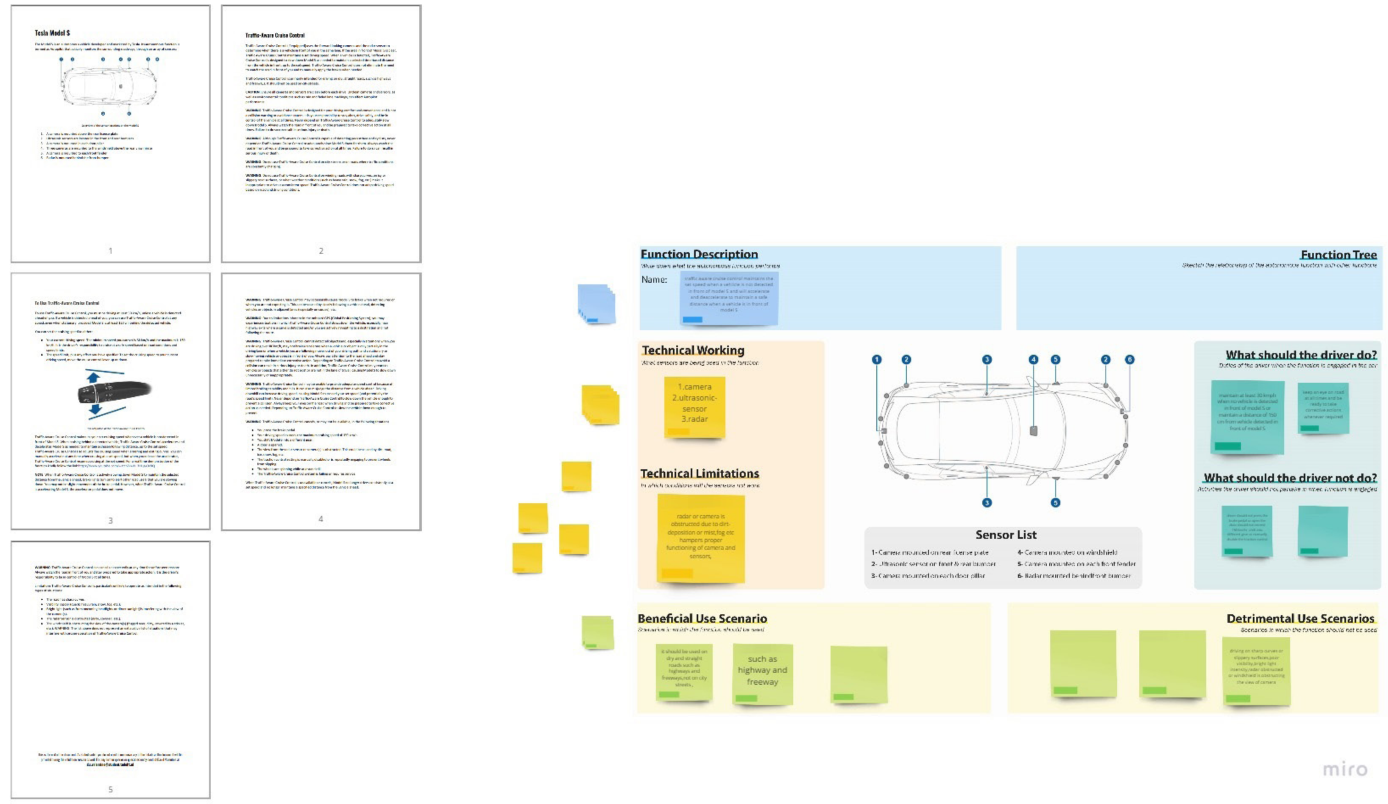


Figure 6.4. A snapshot of a completed testing of the canvas

The validation of the responses of the participants during the session correlated with the insights from the interview. There were however some interesting insights that were seen when comparing the responses. There were some overlaps in the responses of the different questions, this was visible in the case of the section pertaining to "Technical Limitations" and "Detrimental use scenarios". This can be attributed to two reasons, the first is with regard to the interdependency of the two questions and the second is related to the set up of the test. This was also highlighted by one of the participants during the interview *"The aspect of scenarios were difficult as I was limited to the information provided, because I did not see the function in action. I think there was a bit of overlap in the answers for the various questions*". In addition there also existed a second difference which pertained to the level of detailing of each of the sections, between the responses of the participants. Other than the two points discussed above, the responses of the participants correlated to the ideal response to a high degree. There was seen to be a clear understanding of the autonomous function in the responses of the participants recorded on the canvas. The participants provided the correct responses to the three sections: technical working, context of use and user behaviour.

Overall, the visualization canvas was well accepted and the participants liked the ability to explain an autonomous function in a fixed framework that was structured and contained the necessary requirements. The act of filling the canvas was also not seen as a cumbersome task and was considered straightforward. Thus, we can conclude that the autonomous function visualization canvas meets the requirements.

User Decision Matrix

Test & Validation Setup

A similar process as the Automation Function Visualizer was performed to test the user decision matrix as employed in the testing of the Autonomous Function Visualization Canvas. A simple scenario was created, the scenario was overtaking a vehicle on a highway/ freeway. For each session the participants were required to fill in the canvas. After that a single context factor was varied such as day to night, sunny to rainy etc. After varying this single factor the participants were required to fill in the canvas again highlighting the difference in their approach when the context changes. Once both the canvas for both the conditions were created, a semi structured interview was conducted to assess the compliance of the design intervention with the requirements. After each test the matrix was iterated upon adjusting the design to the suggestions of the participants. A similar approach was employed as in the case of autonomous function visualizer for selecting participants.

A second round of validation was performed by analysing the response of the participants recorded in the user decision matrix. The recorded responses were compared to a standard response as filled by the researcher, this was done to compare the deviation of the responses from the standard responses. The second reason was to compare if the matrix was able to capture unique differences of the users.

Requirements for Sensitizing Session

- *Building the context of use including the internal and external context*
- *Describe the decision making process while driving*
- *Integrate the with HMI and eHMI systems*
- *Intuitive and easy to use*

Results

The results of the testing were positive, the participants were able to use the scenarios sheet and the user decision matrix to describe their driving behaviour. In the final iteration of the design the participant was able to fill in the user decision matrix without much instruction confirming that the decision matrix was intuitive and easy to use *"Yeah, I think it was intuitive, yeah I think so....I could have filled it without (researcher) you...I think I understood it"*. In the initial iterations the canvas was not much detailed which made it difficult for the participants to intuitively understand what is expected. For this reason examples were added to help out participants.

> *"Yeah, I think it was intuitive, yeah I think so....I could have filled it without (researcher) you...I think I understood it"*
>
> -Participant 1 (Session-3)

The scenario sheet was while not directly validated was also iterated upon. The reason for this was in the initial testing it was seen that not much detail was being captured and while filling in the user decision matrix the participants were making assumptions as seen in this comment *"Like in the night part there are certain configurations that I check and that might come in the scenario sheet when describing the scenario"*. To counter this more explicit sections were created in the scenario sheet which are shown in the final design. In addition it was seen that the visualization of the driving task played an important role as well. A generic visualization raised more questions about the context, also participants commented that there were inaccuracies in the representation. For this reason the visualization of the driving task must be detailed so as not to create too many assumptions.

> *"Like in the night part there are certain configurations that I check and that might come in the scenario sheet when describing the scenario"*
>
> -Participant 1 (Session-1)

The second round of validation, in which the responses of the participants were compared to a standard response found that there was minimal deviation from the standard response. These differences were attributed to two factors, the first is the variation due to nationality of use of vehicles, different countries have varying driving rules or regulations along with social norms (Rasuli & Tsotsos, 2019) that are not explicitly discussed but influence the driving behaviour, this also corroborated to the the base of dispositional trust by Hoff & Bashir (2015) . The second factor is the level of detailing the scenario canvas that governs the assumptions the participants would make when filling the decision matrix. The matrix is successful in recording individual differences in habits that users have when performing the same driving task.

Overall, the testing and validation showed that the user decision matrix works and can be used within the design process. In addition it was intuitive in nature and was able to capture the change in context factors and their influence on the decision making.

Figure 6.5. A snapshot of one of the testing sessions performed on Miro. As seen two different scenarios were used to explore the difference in the user decision with the change in the context

6.2 | Concept Validation

In parallel with the testing and the iteration of the individual parts of the design intervention. The overall concept also needed to be validated. For this experts in the field of autonomous vehicles and trust were interviewed. The interview was in the form of a presentation where the overall concept was explained and then each design intervention was discussed further in detail. The validation was done in two parts, the first a discussion, in which the expert was encouraged to ask questions about the intervention and have a discussion. The second part was again a semi-structured interview in which the requirements of the design intervention were assessed. Since, the validation of the concept had a little overlap with the testing, some of the insights from the validation session were also incorporated into the iterations and tested subsequently. In total 5 interviews were conducted with experts in various fields that are associated with autonomous vehicles, including designers, engineers and developers.

Requirements for Concept Validation

- *Overall coherence of the toolkit*
- *Perceived value of the toolkit with regard to calibration of trust*
- *Clarity of the toolkit*

Result

Sensitizing Session

The sensitizing session was considered as a good concept as they believe that it allowed for getting the complete development team together and get a common understanding of calibration of trust. It was also suggested that it could lead to other vehicle systems to support or suggest insights with regard to the calibration of trust. A common theme with the majority of the experts was that they would be able to provide more concrete suggestions and insight into the design if they were able to take part in the sessions, but also agreed that it would take more time. The conclusion of the sensitizing session with a trust goal was seen as a good point to end the session as it created a common goal within the development team.

> *“ I fully I fully agree with the sensitizing bit, it’s very important. And also it’s super brilliant as a way of having like, the designers within the development team to get the same goals because that will minimize the..that will minimize no net that’s the word deliberate designs that the user will encounter in the end, which in turn, will give a more appropriate level of trust. So I think it’s a brilliant idea. Very good idea.”*
>
> *-Expert in Trust in Automation (Expert-1)*

Autonomous Function Visualization Canvas

The response to the autonomous function visualization canvas was also positive. The experts agreed that the system approach of representing information for various autonomous systems allowed for a more coherent understanding of the system. They however, did provide some suggestions for improvements. The first was to consider the exact location of using the autonomous function visualization canvas within the development process. This would make it easier for the autonomous technology development team to use it in an effective manner. A second suggestion which was also pointed out in the user testing was the inclusion of additional information such as the working of the sensors and their connection details. However, they also agreed upon the fact that this could detract from the main aim of the design and it is difficult to identify how much information is too much information. Overall, the design was well accepted and the experts believe that there is possibility of evolving the design to add other information in the future.

> *“ I would like to see something like this would be standardized for car companies to to consider all the time when they develop automated vehicles in order to have because then the user always know how to get the information and how to understand the information if it’s equal for all car companies.”*
>
> *-Expert in Trust in Automation (Expert-1)*

User Decision Matrix

The user decision matrix was considered as an important part of the toolkit. They agreed that the ability to break down the decision making of a user in a certain context would aid greatly in the design and development process of the autonomous systems. One suggestion was provided with regard to the scenario sheet as they stated that more information can be added to the scenario sheet. It was pointed out that with more testing it would be possible to identify other categories of information that are missing and can be added to the scenario sheet. With regard to the matrix itself it was seen that the illustrative example as taken from the testing highlighted the difference in driving scenarios as one expert found some of his personal driving behavior different from that of the participant. This again highlights the ability of the user decision matrix to capture personal differences while driving might be missed. It was also suggested that there could be additional use cases for the user decision matrix such as comparing the fully autonomous vehicle with manual driving and looking for differences.

> *“ the context part, I think it’s a very important part. I think that’s it. Really good, unless you said to maybe try a scenario and then see change the context and see what’s happening. I think that can be a really key thing, and also to see. But also to use this for maybe not the user, but only use it for automated system as well. And to see, like, this is what’s happening”*
>
> *-Expert in Trust in Automation (Expert-5)*

Trust Enhancing Communication

The trust enhancing communication was seen as a concise but effective method to convey the important factors that govern trust. It was pointed out that the three principles were relevant to the current research and design approach in the autonomous vehicle area, especially transparency. On the contrary, it was also stated that transparency was a double edged sword,it is important to have transparency but challenging to achieve complete transparency. To overcome this it was pointed out to clarify the expectations of transparency in TEC. One expert believed that TEC should always be considered when communicating about autonomous vehicles and must not be a toolkit but more like a guideline for everybody associated with autonomous vehicles to follow. One improvement that was suggested is with regard to the relation between TEC and the other parts of the toolkit and the role it played in the overall trust formulation. This insight was addressed by improving upon the role of TEC within the product development process and refining its role.

> *“ But in my in my understanding how I would if I was to use it, however easy it was to keep this always in the background like this is the basis of that’s always going to be in my mind”*
>
> *-Expert in Vulnerable Road users (Expert-4)*

Conclusion

Overall coherence of toolkit: The toolkit was overall considered coherent, the only suggestion was to be more specific about the role of TEC within the toolkit. The importance of a sensitizing session at the beginning of the product development process was considered as an important starting point.

Perceived Value: There was seen to be value in the toolkit especially with regard to collaborating for designing for trust among multiple stakeholders. The autonomous function visualization canvas and the user decision matrix allowed for creating common mental models of using autonomous vehicles and also testing out assumptions

> *“ So I think that and given the very multi stakeholder nature of the domain that we work in, I think that this offers a lot of potential in in bringing bringing people together on the same page and streamlining the design process.”*
>
> *-Expert in Vulnerable Road users (Expert-4)*

Clarity of the toolkit: From the interviews with the experts and the tests conducted, the toolkit was seen to be clear to use and the end goal of the various parts of the toolkit were clear. Certain canvases within the sensitizing session were seen to provide some confusion and thus there is a need for support of the facilitator to intervene. The suggestion of identifying concrete intevention points for using the toolkit within the product development process as seen as a point of improvemnt that would make the use of the overall toolkit much simpler and straighforward.

Part 3 | Chapter 7

Project Reflection

The last chapter briefly describes the limitations of the project along with areas of future research. A section with respect to contribution of the project towards autonomous vehicles development has also been added. The report is concluded with a personal reflection on the project

7.1 | Limitations

The project was successfully completed in the stipulated time and the initial design goal of creating a design toolkit aided in designing for calibrated trust was met. There were still some limitations that were present throughout the project. In this section we explore certain limitations:

Context Limitations

- The project was not associated with any organisation developing an autonomous vehicle, which would make the outcome of the project generic and thus additional research would need to be considered to modify the toolkit to meet the needs of a specific organisation.
- The stakeholders described within the project are based on literature and the limited information available on company websites with regard to their autonomous division. Thus, there exists the possibility that some stakeholders might have been overlooked. Also, a lack of understanding of the team structure can have an influence of deployment of the toolkit.

Literature Research Limitations

- Research in the area of mental models was seen as a limiting factor as it is seen that the creation of mental models of using autonomous systems plays an important role in the calibration of trust.
- In the category of road user the focus was on the vulnerable road users and the driver of the autonomous vehicle, no research was performed in the context other vehicles present.
- Product development process selected from the project was based on literature present with regard to designing for automobiles. There is no literature detailing the process of developing autonomous vehicles. Thus, there might be some differences within the prescribed stages within the product development process and the relation between the various systems present within an automobile.
- The research in the field of calibration of trust is an ongoing process and there is certain research that was published during the course of the project that was not incorporated into the literature review.

Qualitative Research Limitations

- The 16 interviews conducted during the qualitative research process needed to be analysed within a stipulated amount of time. While the process of analysis was followed step by step there still remains the possibility to improve upon the analysis and triangulation of the final results

Toolkit Limitations

- The toolkit was designed and tested with a limited number of participants, however more testing can be performed for further iteration on the design.
- The toolkit was not tested in complete but in parts, for the complete validation of the toolkit it will need to be deployed in an actual product development process and then validated for the initial design requirements.
- During the testing of the toolkit limited scenarios were considered, further testing can use diverse scenarios to test the reliability of the toolkit especially the user decision matrix.
- The final validation was conducted by 5 experts, more validation interviews can be conducted with experts in other fields associated with the development of autonomous vehicles such as human factors, data science and software engineers.
- The toolkit was not implemented within an actual product development process. It should thus be the next step to inculcate the toolkit into the product development process of an autonomous vehicle to study if there is a need for additional iterations or modifications.

7.2 | Future Research

- **Naturalistic Studies of Evolving user behavior:** From the literature review and the interview insights it is seen that user behavior is bound to change as one autonomous vehicle is introduced into our society. While simulation of AV interaction provides a glimpse into this interaction there is a need to increase the number of naturalistic studies with regard to AV to understand latent behavioural changes and in some cases emergent behavior. The AV interaction is not only associated with the driver and an Av, but also vulnerable road users and other road users as well.

- **Role of branding on calibration of trust:** Branding plays a major role in the initial mental model formation of consumers about a product. The same is true for AV. This opens an interesting avenue to explore the role of branding, especially innovation driven branding, which is the case of companies such as Waymo and Zoox and how this influences the trust formation and calibration in users.

- **Communicating autonomous capabilities to end users:** Communication was identified as an important aspect of calibration of trust, while the project focused on communication within the product development process. Communication between the company and the user is also important and is an interesting research area to look into the effect different modes of communication have on the users trust and mental model

- **Role of dealerships and after sales service in calibration of trust:** With an increase in digital purchase of automobiles and new forms of mobility appearing. The role of a dealership is changing, it is important to explore what would be the role of a dealership in the future of AV and how the sales and service at dealership can influence the users trust in the AV

- **Organisation change towards designing for AV:** The development of AV technology is not just an evolution of automobiles but also mobility. In the current project we focused on the calibration of trust from a technical standpoint of understanding the technology. However, there are other aspects such as legislature and ethics that play an important role. It is thus, important to look at how organisations need to evolve to match with the demands and challenges that AV pose

7.3 | Contribution Practice

The project was conducted at the cross-section of autonomous technology development, trust and strategic design. In this section we explore the contribution of the current project to each of these sections. Trust in AV is a highly researched topic and is considered as one of the main factors governing the acceptance of autonomous vehicles. However, there is limited research in the area of using calibrated trust to design for autonomous vehicles. Nearly all research pertaining to design for calibrated trust was associated with the design of HMI systems.

Autonomous Vehicles

The project explores the trust formation process of a user with an autonomous vehicle, combining literature and insights from experts from diverse fields to build an understanding of trust in automation. A majority of the challenges to AV have been detailed in existing literature but little importance is given to the change in the design process required to address these barriers. The project sheds light on the importance of communication between various stakeholders to build an AV that can be trusted by the user and the importance of a socio-technical approach to design.

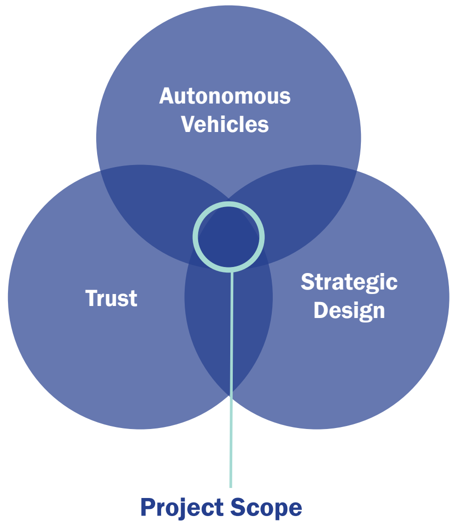
The autonomous function visualization canvas provides an initial design to create templates for communicating autonomous capabilities and standardise the autonomous technology from a more human centered perspective as compared to a technology centric perspective. The user decision matrix can be viewed as an addition to the work of Visser et al. (2016) and Mining et al. (2016) in the approach to provide an approach of understanding the user decision making process when driving an autonomous vehicle.

Calibrated Trust

The project does not take the same direction as other projects focused towards calibration of trust take in the form of focusing on the use phase and more specifically the design of a curated HMI system. Instead it focuses on the product development process. The sensitizing session is one of the first designed sessions for exploring calibrated trust and focusing on a trust goal. The sensitizing session can also be used for other automations and technologies making it a start point to look at workshops and other methods of informing designers and developers about calibrated trust. The project also deviates from the design for learnt trust and looks at situational trust as an aspect that can influence calibration. The project is generally focused on highlighting the need to not only focus on the use phase of the automation/technology but explore the overall product development process if calibration of trust is to be achieved.

Strategic Design

The Calibrated Trust Toolkit used a strategic design approach to designing the final toolkit. Exploring the role of the autonomous technology development team and how to incorporate a socio-technical approach to design by considering the context of use. The project is an initiative to use strategic design to explore addressing the complex design process of an automobile. Pulling back the curtain on the organisational and design changes that are required to meet the needs of future mobility. In this process providing an insight for fellow designers to explore the role of strategic design in the design process of automobiles and similar products and services (aviation, nautical and heavy machinery).



7.4 | Reflection

The project was started three days into the mandatory lockdown imposed by the Dutch government on 16th March. On the 28th August the project will officially end three days before the new academic year starts and partial physical lectures are conducted. This period has truly been a whirlwind of an experience both as a designer and a human being. I have enjoyed the duration of the project and the interaction with my graduation committee. It has been challenging working remotely for nearly 6 months on a single project but reflecting back there have been some important learning points and some inhibitions shattered.

The project requires careful attention to the planning and scheduling of tasks within the stipulated time frame, this involved not only understanding my perceived capabilities but actual capabilities in working from home as compared to working from the faculty. During the literature review and qualitative phase, I encouraged and pushed to improve my research rigor of the project along with working on improving my visualization of data. For research I generally prefer a quantitative approach as it involves less speaking to people and more analysis of data, but the project allowed me to explore my inhibitions of interviewing people and by the end of the project I had conducted and analysed more than 30 interviews.

One of the reasons for selecting the graduation project was to further my association with the field of mobility and explore new technological developments within that field. Reflecting back I am happy to have taken up the project and feel that there is a lot I have learnt about design and research over the duration of the project. My understanding of autonomous technology has improved many folds. Through the project I was able to design for a product development process which I had not done previously. One of the challenging aspects of the project was trust, as described by one of the participants "Trust is an abstract term", which required patience to explore and understand the factors that underlined trust.

The design process and testing process were also insightful as they not only provided insight into the design but also latent points on improving my design process. The meetings with my graduation committee were also helpful in this area allowing me to not only soundboard my assumptions and decisions but also provide critical review of my progress. The interaction with the experts yielded some unexpected outcomes and also advice on conducting the project.

To sum up my learning from the project, the most important learning was with regard to managing the overall design process within a stipulated time frame, as there are certain stages within the design process that are more challenging personally than others. Exploring an abstract concept like trust and grounding it into more tangible and measurable components. Improving my visualization and lastly maintaining research rigor throughout the project. I hope my work inspires others to deep dive further into this topic of adjacent topics.

"Never trust anything that can think for itself if you can't see where it keeps its brain"

J.K. Rowling

Photo by NeONBRAND on Unsplash

Reference

A

Abraham, H., Lee, C., Brady, S., Fitzgerald, C., Mehler, B., Reimer, B., & Coughlin, J. F. (2017, January). Autonomous vehicles and alternatives to driving: trust, preferences, and effects of age. In Proceedings of the transportation research board 96th annual meeting (TRB'17).

Ackermann, C., Beggiato, M., Schubert, S., & Krems, J. F. (2019). An experimental study to investigate design and assessment criteria: What is important for communication between pedestrians and automated vehicles? Applied ergonomics, 75, 272-282.

Adnan, N., Nordin, S. M., bin Bahruddin, M. A., & Ali, M. (2018). How trust can drive forward the user acceptance to the technology? In-vehicle technology for autonomous vehicle. Transportation research part A: policy and practice, 118, 819-836.

Aizenberg, E., & Hoven, J. V. D. (2020). Designing for Human Rights in AI. arXiv preprint arXiv:2005.04949.

Ajzen, I., & Fishbein, M. (1980). Understanding attitudes and predicting social behavior. Upper Saddle River, NJ: Prentice Hall.

Alawadhi, M., Almazrouie, J., Kamil, M., & Khalil, K. A. (2020). Review and analysis of the importance of autonomous vehicles liability: a systematic literature review. International Journal of System Assurance Engineering and Management, 1-23.

B

Bainbridge, L. (1983). Ironies of automation. In Analysis, design and evaluation of man–machine systems (pp. 129-135). Pergamon.

Banks, V. A., Eriksson, A., O'Donoghue, J., & Stanton, N. A.(2018). Is partially automated driving a bad idea? Observations from an on-road study. Applied Ergonomics, 68, 138–145

Bazilinskyy, P., Kyriakidis, M., & de Winter, J. (2015). An international crowdsourcing study into people's statements on fully automated driving. Procedia Manufacturing, 3, 2534-2542.

Bhise, V. D. (2017). Automotive product development: A systems engineering implementation. CRC Press.

Boelhouwer, A., van den Beukel, A. P., van der Voort, M. C., Hottentot, C., de Wit, R. Q., & Martens, M. H. (2020). How are car buyers and car sellers currently informed about ADAS? An investigation among drivers and car sellers in the Netherlands. Transportation Research Interdisciplinary Perspectives, 100103.

Buchler, N., La Fleur, C. G., Hoffman, B., Rajivan, P., Marusich, L., & Lightner, L. (2018). Cyber teaming and role specialization in a cyber security defense competition. Frontiers in psychology, 9, 2133.

C

Calabretta, Gemser & Karpen, G. G. I. (2019). Strategic Design: Eight essential practices every strategic designer must master. BIS.

Casner, S. M., Hutchins, E. L., & Norman, D.(2016). The challenges of partially automated driving. Communications of the ACM, 59(5), 70-77.

Clarke, V., Braun, V., & Hayfield, N. (2015). Thematic analysis. Qualitative psychology: A practical guide to research methods, 222-248.

Craigen, D., Diakun-Thibault, N., & Purse, R. (2014). Defining cybersecurity. Technology Innovation Management Review, 4(10).

D

Davidson, P., & Spinoulas, A. (2015, July). Autonomous vehicles: what could this mean for the future of transport. In Australian Institute of Traffic Planning and Management (AITPM) National Conference, Brisbane, Queensland.

Dawson, J., & Thomson, R. (2018). The future cybersecurity workforce: Going beyond technical skills for successful cyber performance. Frontiers in psychology, 9, 744.

Daziano, R. A., Sarrias, M., & Leard, B. (2017). Are consumers willing to pay to let cars drive for them? Analyzing response to autonomous vehicles. Transportation Research Part C: Emerging Technologies, 78, 150-164.

de Visser, E. J., Cohen, M., Freedy, A., & Parasuraman, R. (2014, June). A design methodology for trust cue calibration in cognitive agents. In International conference on virtual, augmented and mixed reality (pp. 251-262). Springer, Cham.

Dorst, K., & Cross, N. (2001). Creativity in the design process: co-evolution of problem–solution. Design studies, 22(5), 425-437.

Dreibelbis, R. C., Martin, J., Coovert, M. D., & Dorsey, D. W. (2018). The looming cybersecurity crisis and what it means for the practice of industrial and organizational psychology. Industrial and Organizational Psychology, 11(2), 346-365.

E

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

Endsley MR (2017) Autonomous driving systems: a preliminary naturalistic study of theTesla Model S. J Cogn Eng Decis Making 11(3):225–238

Endsley MR (2017) From here to autonomy: lessons learned from human-automation research. Hum Factors 59(1):5–27
Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. Human factors, 37(1), 32-64.

Endsley, M. R. (2019). The limits of highly autonomous vehicles: an uncertain future: Commentary on Hancock (2019) Some pitfalls in the promises of automated and autonomous vehicles. Ergonomics, 62(4), 496-499.

Etzioni, A. (2019). Cyber trust. Journal of Business Ethics, 156(1), 1-13.

Everson, E. (2016). Privacy by design: Taking ctrl of big data. Clev. St. L. Rev., 65, 27.

F

Fagan, M., & Khan, M. M. H. (2018). To follow or not to follow: a study of user motivations around cybersecurity advice. IEEE Internet Computing, 22(5), 25-34.

Fagnant, D. J., & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. Transportation Research Part A: Policy and Practice, 77, 167-181.

Fraedrich, E., & Lenz, B. (2014). Automated driving: Individual and societal aspects. Transportation research record, 2416(1), 64-72.

Franssen, M. (2015). Design for values and operator roles in sociotechnical systems. Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains, 117-149.

Fridman, L. (2018). Human-centered autonomous vehicle systems: Principles of effective shared autonomy. arXiv preprint arXiv:1810.01835.

Friedman, B., & Hendry, D. G. (2019). Value sensitive design: Shaping technology with moral imagination. Mit Press.

Fuller, R., (2005). Towards a general theory of driver behaviour. Accid. Anal. Prev. 37,461–472.

H

Habibovic, A., Lundgren, V. M., Andersson, J., Klingegård, M., Lagström, T., Sirkka, A., ... & Saluäär, D. (2018). Communicating intent of automated vehicles to pedestrians. Frontiers in psychology, 9, 1336.

Haddal, R., & Hayden, N. K. (2018). Autonomous Systems Artificial Intelligence and Safeguards (No. SAND2018-9787C). Sandia National Lab.(SNL-NM), Albuquerque, NM (United States).

Hadlington, L., & Murphy, K. (2018). Is media multitasking good for cybersecurity? Exploring the relationship between media multitasking and everyday cognitive failures on self-reported risky cybersecurity behaviors. Cyberpsychology, Behavior, and Social Networking, 21(3), 168-172.

Henshel, D., Cains, M. G., Hoffman, B., & Kelley, T. (2015). Trust as a human factor in holistic cyber security risk assessment. Procedia Manufacturing, 3, 1117-1124.

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. Human Factors, 57, 407–434

Hoffman, R. R. (2017). A taxonomy of emergent trusting in the human–machine relationship. Cognitive systems engineering: The future for a changing world, 137-163.

J

Johnston, D. (1973). Road accident casualty: A critique of the literature and an illustrative case. Ontario: Grand Rounds. Department of Psychiatry, Hotel Dieu Hospital.

Juang, K., & Greenstein, J. (2018). Integrating visual mnemonics and input feedback with passphrases to improve the usability and security of digital authentication. Human factors, 60(5), 658-668.

K

Kalra, N., & Paddock, S. M. (2016). Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?. Transportation Research Part A: Policy and Practice, 94, 182-193.

L

Langheinrich, M. (2001, September). Privacy by design—principles of privacy-aware ubiquitous systems. In International conference on Ubiquitous Computing (pp. 273-291). Springer, Berlin, Heidelberg.

Lee, J. D., & Kolodge, K. (2020). Exploring trust in self-driving vehicles through text analysis. Human factors, 62(2), 260-277.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors, 46, 50–80.

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

Lim, H. S. M., & Taeihagh, A. (2018). Autonomous vehicles for smart and sustainable cities: An in-depth exploration of privacy and cybersecurity implications. Energies, 11(5), 1062.

Lin R, Ma L, Zhang W (2018) An interview study exploring Tesla driver's behavioral adaptation. Appl Ergon 72:37–47

Linkov, V., Zámečník, P., Havlíčková, D., & Pai, C. W. (2019). Human factors in the cybersecurity of autonomous cars: trends in current research. Frontiers in psychology, 10, 995.

M

Maguire, M., & Delahunt, B. (2017). Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars. All Ireland Journal of Higher Education, 9(3).

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. Academy of Management

Mica R Endsley. 2019. Situation Awareness in Future Autonomous Vehicles: Beware of the Unexpected. In Proceedings ofthe 20th Congress of the International Ergonomics Association (IEA 2018). Springer International Publishing, 303–309.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2013). Research design and management. Qualitative data analysis: A methods sourcebook (3rd ed., pp. 17-54). Thousand Oaks, CA: SAGE.

Mirnig AG, Wintersberger P, Sutter C, Ziegler J. A framework for analyzing and calibrating trust in automated vehicles. In: Adjunct Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications; 2017. p. 33–38. https://doi.org/10.1145/3004323.3004331.

Morley, C. (2019). Automotive Industry Life Cycle - Shorter Timelines. Https://Www.Jabil.Com/Blog/Automotive-Industry-Trends-Point-to-Shorter-Product-Development-Cycles.Html. Retrieved August 19, 2020, from https://www.jabil.com/blog/automotive-industry-trends-point-to-shorter-product-development-cycles.html

N

Noy, I. Y., Shinar, D., & Horrey, W. J. (2018). Automated driving: Safety blind spots. Safety science, 102, 68-78.

O

Okamura, K., & Yamada, S. (2018). Adaptive trust calibration for supervised autonomous vehicles. In Adjunct Proceedings of the10th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications. Retrieved fromhttps://dl.acm.org/citation.cfm?doid=3239092.3265948

P

Parkinson, S., Ward, P., Wilson, K., and Miller, J. (2017). Cyber threats facing autonomous and connected vehicles: future challenges. IEEE Trans. Intel. Transport. Syst. 18, 2898–2915. doi: 10.1109/tits.2017.2665968

Patton, M. Q. (2002). Qualitative interviewing. Qualitative research & evaluation methods (3rd ed., pp. 339-418). Thousand Oaks, CA: SAGE.

Paul R. Carlile, (2002) A Pragmatic View of Knowledge and Boundaries: Boundary Objects in New Product Development. Organization Science 13(4):442-455. https://doi.org/10.1287/orsc.13.4.442.2953

Pype, P.; Daalderop, G.; Schulz-Kamm, E.; Walters, E.; von Grafenstein, M. Privacy and security in autonomous vehicles. In Automated Driving; Springer: Berlin/Heidelberg, Germany, 2017; pp. 17–27.

R

Rangwala, S. (2020, August 1). Amazon's Zoox Acquisition–Is LiDAR Next? Forbes.https://www.forbes.com/sites/sabbirrangwala/2020/08/01/amazon-zoox-acquisition-is-lidar-next/#324784367eda

Rasmussen, J., Pejterson, A. M., & Goodstein, L. P. (1994). Cognitive systems engineering. New York: Wiley

Rasouli, A., & Tsotsos, J. K. (2019). Autonomous vehicles that interact with pedestrians: A survey of theory and practice. IEEE transactions on intelligent transportation systems.

Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. Journal of personality and social psychology, 49(1), 95.

S

Saldaña, J. (2012). First cycle coding methods. The coding manual for qualitative researchers (2nd ed., pp. 58-186). London: SAGE.

Saldaña, J. (2012). Second cycle coding methods. The coding manual for qualitative researchers (2nd ed., pp. 207 245). London: SAGE.

Sanders, E. B.-N., & Stappers, P. J. (2012). Convivial design toolbox: Generative research for the front end of design.

Schoettle, B., & Sivak, M. (2014). A survey of public opinion about autonomous and self-driving vehicles in the US, the UK, and Australia. University of Michigan, Ann Arbor, Transportation Research Institute.

Schön, D. A. (1983). The reflective practitioner: How professionals think in action.

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019, January). Fairness and abstraction in sociotechnical systems. In Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 59-68).

Smith, K. H., & Preiser, W. F. (2001). Universal design handbook. McGraw-Hill.

Sörensen, D. (2006). The automotive development process (pp. 13-53). DUV.

Spencer, L., Ritchie, J., Ormston. R., O'Connor, W. & Barnard, M. (2014). Analysis: Principles and Processes. In J. Ritchie, J, Lewis, C. McNaughton Nicholls & R. Ormston (eds.), Qualitative research practice: A guide for social science students and researchers (2nd ed., pp.269- 293). London: SAGE

T

Templeton, B. (2020, June 2). Tesla In Taiwan Crashes Directly Into Overturned Truck, Ignores Pedestrian, With Autopilot On. Forbes. https://www.forbes.com/sites/bradtempleton/2020/06/02/tesla-in-taiwan-crashes-directly-into-overturned-truck-ignores-pedestrian-with-autopilot-on/#43df487058e5

Tenhundfeld, N. L., de Visser, E. J., Haring, K. S., Ries, A. J., Finomore, V. S., & Tossell, C. C. (2019). Calibrating Trust in Automation Through Familiarity With the Auto Parking Feature of a Tesla Model X. Journal of Cognitive Engineering and Decision Making, 13(4), 279-294.

Tenhundfeld, N. L., de Visser, E. J., Ries, A. J., Finomore, V. S., & Tossell, C. C. (2019). Trust and distrust of automated parking in a Tesla Model X. Human factors, 0018720819865412.

U

Ulrich, K., Eppinger, S., & Yang, M. C. (2019). Product Design and Development (7th ed., Vol. 1). McGraw-Hill Education.
van de Poel, I. (2018). Design for value change. Ethics and Information Technology, 1-5.

W

Wang, W., Zhou, F., Li, W., & Budd, J. (2018). Designing the product-service system for autonomous vehicles. IT Professional, 20(6), 62-69.

Weber, M. (2014). Where to? a history of autonomous vehicles. Computer History Museum, 8.

Whitworth, B., & De Moor, A. (Eds.). (2009). Handbook of research on socio-technical design and social networking systems. IGI Global.

Wicker, S. B., & Schrader, D. E. (2010). Privacy-aware design principles for information networks. Proceedings of the IEEE, 99(2), 330-350.

Wong, Y. Z., Hensher, D. A., & Mulley, C. (2018). Emerging transport technologies and the modal efficiency framework: A case for mobility as a service (MaaS).

Y

Yang, J. H., Han, J., & Park, J. M. (2017, May). Toward Defining Driving Automation from a Human-Centered Perspective. In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (pp. 2248-2254).

Design for Calibrated Trust for Acceptance of Autonomous Vehicles