UNIVERSITY OF GOTHENBURG

Autonomous cars and agency:

An empirical study on the coexistence of artificial drivers and humans in traffic

Självkörande bilar och agency:

En empirisk studie om samexistensen mellan artificiella förare och människor i trafiken

Linus Boström
Eric Ohlsson

Master of Science Thesis in Informatics

Report nr. 2017:052

University of Gothenburg
Department of Applied Information Technology
Gothenburg, Sweden, May 2017

Abstract

As autonomous cars are introduced in the social environment of traffic it is uncertain how they will enact agency. Prior research has focused on solely technical aspects, overlooking the social. The aim was thus to investigate how agency is attributed to autonomous cars in traffic. Using an interpretative qualitative research approach, we explored how eight stakeholders from various domains attributed agency to autonomous cars. This was accomplished by having them solve scenarios that highlights agential issues with autonomous cars-human interaction. The results showed that most endeavors of introducing autonomous cars involved adapting the environment to the technology, something that is problematic both in theory and practice. Some respondents from academia attributed human-like agency to autonomous cars as they with programmed aggressive behaviour were to actively cooperate with humans. Nevertheless, this view proved to be in minority as practitioners and respondents from public sector attributed the pre existing Information Systems (IS) definition of agency where technology acts passively under human control. This was also reflected by technology in use as well as rules that govern how autonomous cars operate in the real world, prompting us to favor the IS definition. However, the autonomous cars were at times not under *direct* human control since they under highly constricted conditions operated autonomously, making the IS definition somewhat inadequate. Consequently, we coined a new definition called "*limited autonomous agency*" that more adequately reflects how autonomous cars operates autonomously while being in arm's reach of humans.

Keywords: autonomy, agency, autonomous agency, autonomous technology, autonomous cars

Abstrakt

Det är osäkert hur självkörande bilar kommer utöva agency när de introduceras i trafikens sociala miljö. Forskning inom området har huvudsakligen adresserat tekniska aspekter och således förbisett de sociala. Syftet med studien var därför att undersöka hur agency attribueras till självkörande bilar när de ska interagera med människor i trafiken. Genom att använda en tolkande kvalitativ ansats undersöktes hur åtta intressenter från olika domäner attribuerade agency till självkörande bilar. Detta genomfördes genom att vi lät dem lösa scenarier som påvisar agencyproblematik gällande interaktionen mellan självkörande bilar och människor. Resultatet påvisade att den framträdande metoden för att införa självkörande bilar innebär att omgivningen anpassas till tekniken, något som är problematiskt både i teorin och praktiken. Vissa respondenter från akademin tilldelade bilarna en form av mänsklig agency då de aktivt kunde samarbeta med människor genom programmerat aggressivt beteende. Denna uppfattning visade sig emellertid vara i minoritet då både praktiker och respondenter från offentlig sektor tillämpade den vedertagna informatikdefinitionen av agency där teknik passivt agerar under mänsklig kontroll. Den senare uppfattningen återspeglades av teknik i användning samt regler som styr hur självkörande bilar skall fungera i praktiken. Detta fick oss att identifiera informatikdefinitionen som mest framträdande. Ibland var dock de självkörande bilarna inte under *direkt* mänsklig kontroll då de under strikta omständigheter verkade autonomt. Informatikdefinitionen var därför inte helt adekvat vilket ledde oss till att föreslå en ny definition kallad "*begränsad autonom agency*". Denna definition speglar hur självkörande bilar fungerar autonomt samtidigt som de befinner sig inom en armlängds avstånd från mänsklig kontroll.

Nyckelord: autonomi, agency, autonom agency, autonom teknik, självkörande bilar

Acknowledgements

We would like to start off by thanking our supervisor Ioanna Constantiou for guiding us on the at times tricky roads of academic research. Ioanna gave us great feedback and steered us back on track when needed.

We also give thanks to all respondents who participated in the study for taking the time to share their valuable knowledge with us.

Thank you!

Linus Boström and Eric Ohlsson
2017-05-24

Table of contents

1. Introduction

Autonomous cars are on the verge of becoming a reality on public roads, where many large car manufacturers are investing heavily in research and development of autonomous driving (cf. Brewster 2016; Gomes 2014; Kennedy 2016; Tesla 2017; Volvo 2017).

Autonomy in cars have evolved from managing mundane tasks of routine operations towards taking full control to the car itself (cf. Anderson et al. 2016; Bengler et al. 2014; NHTSA 2013; SAE J3016 2016). Some of these cars are still relying on human intervention in the event of unforeseen events (Brown & Laurier 2017), while others have the higher level of automation where the system is able to operate completely on its own (Volvo 2017). This would correspond to level 4 or 5 on SAE International's SAE J3016, the de facto standard for classifying vehicle automation (SAE J3016 2016). This study focuses on level 4 or 5 in this standard, where control over the vehicle is handled by the system alone without the need of direct human fallback.

Reasons why autonomous driving technology is being pushed forward often transcend speculation. Notions of increased productivity where the driver can perform work instead of driving the car is often mentioned, as well as increased quality of life since drivers can interact more with the family while being on the road (cf. Anderson et al. 2016; Holmberg et al. 2016; Tesla 2017; Volvo 2017). It is also proposed autonomous driving can assist disabled people better than regular cars (Anderson et al. 2016; Winner & Wachenfeld 2016) as well as being more friendly to the environment, often through mobility as a service (MaaS) solutions (cf. Fraedrich et al. 2015; Holmberg et al. 2016; Johansson 2017; Näringsdepartementet 2016; Schoettle & Sivak 2015).

Noteworthy is the amount of support given to autonomous driving from governments, where laws are being passed to get autonomous vehicles operable on the roads quickly (Kang 2016; Näringsdepartementet 2016). Reasons given are often prospects of increased security by mitigating the risks associated with human control over vehicles (Anderson et al. 2016; Bengler et al. 2014; Näringsdepartementet 2016).

As mentioned above, what the future might bring with autonomous driving is uncertain. The timeframe for adopting autonomous driving is heavily debated as well (Beiker 2014). Fraedrich et al. (2015) claim that it is highly unlikely that a unison shift to fully autonomous cars will occur where regular cars are banned over night. They thus call for research in settings where mixed traffic occur. In such a scenario, autonomous and regular cars will occupy the same streets. The same authors mention that research have mostly been done on technical aspects, overlooking the social.

Similarly, in the field of Human-Robotic Interaction (HRI), where autonomous vehicles have been of interest for a substantial period of time, Sheridan (2016) identifies the social aspect of autonomous driving as an area that demands much further research. Drawing on the previous work by Gao et al. (2006), Lee and See (2004), Schoettle and Sivak (2015) and Seppelt and Lee (2007), Sheridan (2016, p. 528) states that:

> *"It is becoming clear that many complex traffic situations are exceedingly difficult for computer vision and artificial intelligence to "understand" and that many accidents are avoided by social interaction between drivers, such as mutual eye contact, hand signals and so on."*

Consequently, this study identifies that mixing autonomous cars with humans will be the most likely way forwards. We thus propose that it is important in seeing roads as social systems where different road users need to interact and cooperate (Brown & Laurier 2017; Fraedrich et al. 2015; Juhlin 2010). In this sense, Juhlin (2010, p. 58) argues how important it is for the autonomous cars to be socially competent:

> *"Computers, running by rules or algorithms, must function together with other road users. They must adapt to them, or the drivers will have to adapt to the new machines. If the artificial drivers are socially incompetent, this could put serious strains on other road users."*

Brown and Laurier (2017) highlights in a similar manner that the social environment of traffic requires drivers to interact with others. This is illustrated in their study where they have observed many hours of in footage of autonomous cars driving in regular traffic. In several occasions, the autonomous cars inadvertently signal unwanted signals to human road occupants with the car's own "body language". For instance, the Google car while being programmed as being careful, comes off to other drivers as a slow and hesitant driver which may invoke certain behaviour of other road users. In another instance, they identify that the robotic coordination of a Tesla car is being recognized as rude as it fails to understand when another car wants to enter its lane. The main point however, is that autonomous cars have issues in dealing with intentionality when introduced in the embodied social environment of traffic.

To put all of this in theoretical terms, problems occur when technology, that is, autonomous cars, is to act on own volition in social environments. The examples of Brown and Laurier (2017) indicates that current autonomous cars has difficulties to show and understand intentions, traits often considered part of enacting human agency (Leonardi 2012; Rose et al. 2005). And in effect, how do we get autonomous cars to work in this context that is reliant on cooperation to function? With the current advancements in computer science and technology, it is becoming increasingly possible to simulate human traits (Rose et al. 2005; Russel & Norvig 2016). The question is whether technology should mimic the behaviour of humans that occupies the environment it is introduced into. That is, should autonomous cars itself enact human-like agency by understanding and showing intentionality, or should the introduction of autonomous cars in traffic be managed in other ways?

The ability to understand and show intentionality have so far been reserved for humans in the field of Information System (IS), where technology mainly have been seen as "tools" for humans to use (Andersen et al. 2016). As such, technological agency have been described as being instigated by humans (cf. Leonardi 2012; Orlikowski 2010; Pickering 1993). In other words, technology has been seen as reliant on humans to function.

In the scope of this study however, technology is seemingly to operate in a social environment of traffic absent from direct human control. As such, it is arguably not to be used as a tool, but should act on own volition. This proposes that the previous mentions of technological agency might need to be revised in order to encompass autonomous phenomena where agency is not instigated by human action (Andersen et al. 2016). This does not mean that the current theories of agency in the field of IS are invalid, merely that they have been used to describe and understand non-autonomous technology.

As such, describing traffic as a social environment and based on previous theories of agency, the study aims to explore how agency is attributed to autonomous cars. This is achieved by interviewing researchers, policymakers and practitioners, all involved in autonomous driving. The interviews explores four traffic scenarios created by the authors. The scenarios pinpoints agential issues that arise when autonomous cars are introduced in traffic, more specifically how technology can understand and show intent when interacting with humans. This will provide an understanding how agency can be attributed to autonomous cars that is to operate in a highly contingent social environment. As such, it will provide insight into if and how the current theories of agency is adequate to describe these new phenomena as well as exploring how autonomous cars can be introduced in the social environment of traffic. The research question thus reads:

How is agency attributed to autonomous cars?

The paper has two audiences. First, it responds to the call of research for the theoretical field of agency in IS, where autonomous agency is largely unexplored (Andersen et al. 2016). Secondly, it answers the call of research of the social interaction of autonomous cars in traffic (Fraedrich et al. 2015; Sheridan 2016). The study should thus be of interest to both researchers as well as practitioners involved in the development and implementation of autonomous cars.

The study is limited to autonomous driving in personal vehicles. Furthermore, the study does not include or take into consideration the socio-economic implications often discussed in relation to autonomous cars. Lastly, the study takes the position that human intelligence is different from technological intelligence (Floridi 2014). However, we will not enter the debate whether or not technology will possess "true" intelligence.

1.1 Organization of the paper

The paper is organized as follows:

The chapter “Theoretical background” depicts theories on autonomous cars and agency relevant to the scope of research. More specifically, an initial state of the art of autonomous cars, how they are able to operate and how they can be classified under different levels of vehicle automation. Traffic is then described as a social environment that relies on cooperation to function. Further, human and technological agency is addressed and how these relate to autonomous cars.

The chapter “Research approach” contains the research method where we describe what type of research method that were used as well as arguments of why we decided to use it. Moreover, a concise description of the respondents that participated are presented along with the traffic scenarios that were used during the interviews.

The chapter “Empirical findings” presents the empirical data under the themes identified in the data analysis phase. The data is presented with representative quotes and natural text that points out similarities and contrasting views from the respondents.

The chapter “Discussion” addresses first how autonomous cars can be introduced in the social environment of traffic followed by possible explanations how agency can be attributed to autonomous cars. Similarities and contrasting views found in the empirical data are discussed and related to prior theory and other relevant literature. We also state contributions to theory and practice, limitations as well as identifying areas for further research.

Lastly, in the chapter “Conclusions”, we sum up our findings.

2. Theoretical background

This section aims to give an understanding of the concepts at hand and how they help in answering our research question. Initially, we present a state of the art of autonomous cars, how they are able to operate and subsequently what classification of autonomy that is relevant for this study. Following, traffic is described as a social environment that affords cooperation between actors in it, something we argue must be taken into consideration as autonomous cars are to coexist with humans in traffic. Further, the relationship between human and technological agency is described. The last section addresses the interplay between agents as well as placing it in the context of traffic as a social environment

2.1 State of the art of autonomous driving

At heart of autonomous driving is the notion of moving control from the human to the technology. Recently, as technology have matured to permit more advanced functions, it has become possible to move control towards technology alone. In other words, it is possible for autonomous cars to be autonomous in a true sense. Therefore, it raises questions if one should give full control to the autonomous car and what impacts it in doing so could have. This section gives a brief introduction to how autonomous cars function in order to understand the locus of study.

An early stage of autonomy can be seen in the transition from horse carriages to automobiles as horses at times would undertake autonomous missions as they brought a carriage home safely even if the driver was not fit enough for the journey (Maurer 2016). However, the story of autonomous cars began in the United States at the beginning of the twentieth century due to the sharp rise of traffic accidents, caused by human errors that eventually led to ideas to substitute error-prone humans with technology (Kröger 2016).

What is apparent is that autonomous driving has been an area of research for many decades, first by academic and then later on by industry (Anderson et al. 2016; Kröger 2016). In the last decade, there has been immense strides in research regarding autonomous driving on complex roads (Villasenor 2014). In recent years, the Defense Advanced Research Projects Agency, better known as DARPA, organized a series of competitions between 2003 and 2007, where autonomous cars had to navigate safely on roads with respect to other robots, human drivers and the environment without any input from humans (Anderson et al. 2016; Bengler et al. 2014; Brown & Laurier 2017; Kröger 2016; Sheridan 2016). These competitions broadly accelerated advancements in the technology of autonomous vehicles and broadened the scope of what can be established with technology (Anderson et al. 2016).

Autonomous cars sense the environment through sensors, such as cameras, radars and lidars (Anderson et al. 2016; Becker et al. 2017). Their decision making are often based on a fusion of gathered environmental data and previous information (Anderson et al. 2016; Leitner et al. 2017). Decisions are mainly made through the use of hierarchical finite state machines, where high level goals, such as taking the correct route, is superordinate to routine tasks, such as steering or other types of basic control over vehicular functions (Kurt & Özgüner 2013). However, it is not entirely clear how the technology is to be used to understand humans, especially as the role of artificial intelligence (AI) is not fully understood in relation to autonomous cars (Kaznov et al. 2017; Sheridan 2016).

Autonomous cars in general share many characteristics, but their application have differed somewhat between manufacturers, making it difficult to distinguish what is really meant when the term autonomy is used to describe a car. In order to clarify this, standards with different levels of automation have been developed. It is important to highlight the differences between these levels as they signify various levels of human involvement in the operation of autonomous cars and in effect how autonomous they really are. These are presented in the following section.

2.2 Levels of vehicle automation

As many concepts have been used to describe automated vehicles, such as "autonomous", "driverless", and "self-driving", different initiatives came underway in the mid 2000 in order to define different levels of automation (Beiker 2014). Two definitions of automation standards are prevalent; the National Highway Traffic Safety Administration, often referred to as NHTSA (NHTSA 2013) and SAE J3016 (SAE J3016 2016). The latter standard, however, is considered more consistent with industry practice as well as being less ambiguous than the former standard, it has therefore even been adopted by NHTSA itself recently (NHTSA 2016). We will thus use the SAE J3016 standard for describing autonomous cars in this study.

At level 0, the driver is at all times in complete operational control even when enhanced by warning or intervention systems. At level 1, the driver can be assisted by a driving automation system of either steering or acceleration/deceleration, but the driver still has overall control of the vehicle. At level 2, the driver can be assisted by one or more driving automation systems of both steering and acceleration/deceleration, but the driver still has overall control of the vehicle. At level 3, the car can operate by itself with the expectation that the human driver will respond to a request to intervene if needed. At level 4, the car is able to operate by itself even if the human driver does not respond to a request to intervene. At level 5, the car is able to operate by itself under all circumstances (SAE J3016 2016).

To put this in context, this study focuses on level 4 or 5 of automation where the car supposedly is to handle all situations without direct human fallback. As such, the car should be able to understand its environment at all times. But what is the environment it is to operate in more precisely?

2.3 Traffic as a social environment

As autonomous cars are to hit our streets, we find it important to address how these streets actually function in the real world. As such, we identify roads and subsequently traffic as very complex environments that are highly contingent in nature. Adding to this is that traffic consists of human beings that further adds to the complexity. These arguments are elaborated below.

Traffic can as mentioned previously be seen as social environments which are to be used by different types of actors, such as cars, buses, cyclists and pedestrians at the same time (Juhlin 2010). In order to avoid accidents and disturbances, these actors have to interact and cooperate with each other (Brown & Laurier 2017; Fraedrich et al. 2015; Juhlin 2010; Sheridan 2016).

A set of formal traffic rules, such as driving on the right hand side or follow the speed limit makes the interaction between drivers and other actors on the road easier as these rules have to be followed (Juhlin 2010). In contrast, and this is where it gets problematic, is that human drivers and pedestrians use informal rules naturally and intuitively. For instance, pedestrians that intend to cross a street might use eye contact to ensure an approaching driver has seen them and a driver can use hand signals or body language to signal others what he or she is about to do or wants others to do (Färber 2016). These examples of informal rules are much harder for autonomous cars to understand (Sheridan 2016). However, informal rules go beyond mere eye contact, hand signals and body language, as humans rely on both courtesy and intuition to cooperate in traffic (Brown & Laurier 2017; Fraedrich et al. 2015; Juhlin 2010). Moreover, informal rules can also be "flow priority", where formal rules can be bent in order to increase the flow of traffic. In a practical sense, this could be speeding up to let another car in from a junction or letting a car with more momentum pass even though it does not have the legal right to do so (Juhlin 2010).

Adding autonomous cars to the highly contingent social environment of traffic puts more to the point that previously, only humans have interacted with each other. Now, technology is to interact with humans in this environment seemingly on its own devices. As such, we can ask how much control the technology is to have and what decisions it can and should take? Can it understand and cooperate with humans autonomously? To put it in context of our study, traffic has previously housed only human agents, now, this same environment is also to house technological agents in the form of autonomous cars. To understand this interplay, we address agency theories in the following section below.

2.4 Human and material agency

As previously discussed, traffic has mainly housed human agents. Someone or something that enters and acts in this environment should thus arguably understand humans (Juhlin 2010; Sheridan 2016). Now, traffic is to house both humans and autonomous cars (Fraedrich et al. 2015) which has proved troublesome (Brown & Laurier 2017) as humans and technology are seemingly different (Floridi 2014; Rose et al. 2005). This section addresses these differences using the concept of agency.

What constitutes agency have been widely discussed over the years where a main discussion in the field of IS have focused on if and how technology influences human behaviour (cf. Leonardi 2012; Orlikowski 2010). What agency actually is, however, widely differs (Leonardi 2012), and to give some clarity into this debate, we must first understand how the concept is approached from different angles.

At a basic level, we can differentiate agency between that of humans and technology (Leonardi 2012; Pickering 1993). The classification of technology is not that simple either as the former states, we will however not go into that debate in this paper, but we can mention that material agency could be differentiated from natural objects and objects that humans create, e.g. technology (Pickering 1993). We will focus on the technological aspects since the paper is dealing with artifacts created by humans. That is, autonomous cars.

Pickering (1993) argues that material agency is different from human agency in that it does not contain any intent from which an action is taken. Leonardi (2012) further develops this idea and states that material agency is the product of human agency since humans uses technology to fulfil their intent. This view seems to be common in the field of IS, where Kaptelinin and Nardi (2006) also state that human agency is a precursor for material agency. The latter explains that human agency is superordinate to material agency since humans creates the machines that is used to carry out their will. In this view, technology has no will to operate on its own but is an instrument of human intentionality. And as such, is autonomous cars still an instrument of human intentionality when it is to act on its own devices?

In AI research, an agent is often considered anything that senses its environment through sensors and acts on that information through devices (Russel & Norvig 2016). These agents are considered autonomous since they do not require direct human manipulation to function, meaning that they can perform actions solely on the basis on their environment as well as prior and accumulated knowledge. Agents within AI are however given very specific tasks which they are to solve (ibid). In the context of autonomous cars, they would be given very strict rules they are to follow to be classified under this definition. Autonomous agents have been addressed similarly in the field of computer science as systems that react to complex stimuli through previous design (Brustoloni 1991; Maes 1995). In these cases, an autonomous agent has agency in the scope of its design made by humans, similar to the notion of technological agency in the field of IS research. However, this entails that the agent has a clear goal it is to fulfil.

Rose et al. (2005) claim that as technology moves closer to automation, the previous arguments where technology merely extends human agency becomes harder to make. This can be as technology enters domains that may be contingent to the point that is it not possible to foresee and program for every possible action or situation it will encounter. We would agree, since on top of that, technology have a reputation of not being used as intended or being as stable as it was designed (Orlikowski & Iacono 2001). On the same note, in seeing traffic as a highly contingent social environment, it is dubious whether one could foresee every situation an autonomous car is to encounter.

Moreover, as we are moving towards autonomy, Andersen et al. (2016) argue that agency theories in IS research have so far been conceptualized from the position of human agents. That is, technology have inherently been treated as a tool in understanding the interplay between humans and technology. As we are closing in on autonomy, agency can be seen as being transferred towards the artifact alone, meaning that it is not to be regarded as a tool anymore but rather something that can act alone without any human involvement at all. Again, this points towards the scope of the study, where it is unclear whether autonomous cars should be seen as tools instigated by humans or acting on own volition.

This gets further problematic since technology, apart from lacking intentionality, do not share human characteristics. Rose et al. (2005, p. 14) state that "*self awareness, social awareness, interpretation, intentionality and the attribution of agency to others*" are inherently human traits that are not possible for technology to inhibit. They however leave the future open for speculation, saying that these traits can increasingly be simulated by programming. And in autonomous cars, can we program human traits, and if so, should we program them?

2.5 Towards agency in autonomous cars

So far, we have touched upon agency in separate forms, but what is also important is the interplay between different agents as we identify autonomous cars to operate in a social environment, where cooperation have been key in enacting traffic (Juhlin 2010).

When humans cooperate to pursue common goals, this can be referred to as social agency (Leonardi 2012; Pickering 2001). Technology have often been seen as a mediator of action (Orlikowski 2007) or as being imbricated or intertwined (Leonardi 2012) with its use. We can acknowledge that technology can be seen this way, but in the scope of this study, the focus does not lie on exploring the sociomaterial entanglement of humans and autonomous cars. Instead, it deals with if and how technology is to enact agency at the same levels as humans. As discussed before, agency has mainly circulated around the axiom that humans are the instigators of action. Now, we are instead moving towards a situation where technology is to act on own volition in human environments. This does not exclude that technology is entangled with its use, it merely provides a different perspective of who is considered the instigator of agency. It does not render these theories incomplete, it puts them in a different domain.

As discussed in the previous section, we identify traffic as a social environment where different actors must cooperate (Fraedrich et al. 2015; Juhlin 2010; Sheridan 2016). We would thus argue that agents in traffic in fact are enacting social agency as they do so by understanding others, making themselves understood and cooperating to improve informal rules, such as flow. Humans do this effortlessly since the enactment is a product of human traits and human nature (Floridi 2014; Rose et al. 2005). But how is this addressed in the area of autonomous cars when they are to operate in the same social environment as humans that relies on both the formal rules as well as the informal rules that are constantly enacted through cooperation? Should they imitate or simulate human behaviour? Are they to be presented different than humans, or are they to "melt in" so one could not tell if it is a human or an artificial driver?

In order to understand how this interplay can be addressed, we propose four traffic scenarios (see section 3.3) that raises questions how agency is attributed to autonomous cars. Based on how the different scenarios are "solved" or reasoned about, we can draw conclusions how different respondents are attributing agency to autonomous cars when it is to operate in a real world social environment. The scenarios stresses the questions of intentionality as previously mentioned, if the car itself is to deal with complex interactions with humans or if these scenarios are to be managed in other ways. This will give a multidimensional view of how agency is attributed to autonomous cars that is to operate in traffic.

3. Research approach

As previously discussed, autonomous driving in general and its relation to social issues in particular is largely new and unexplored. As such, we found a qualitative empirical approach appropriate as it is considered suitable for these conditions (Klein & Myers 1999; Myers 1997; Walsham 2014). Moreover, since we are to study something that is at the moment non-existing or rather soon to be implemented, we found that we are to deal with perceptions of how new technology could or should be used. As such, we adhered to an interpretive (Walsham 2006) mindset, where we aimed to get a holistic view of the focus of study by interviewing different respondents.

Since the field of study is quite new and unexplored, we further broadened our literature search from the field of IS into other domains, such as Human-Computer Interaction (HCI) and Human-Robot Interaction (HRI), this approach could also be deemed appropriate since IS is an interdisciplinary field (Webster & Watson 2002).

3.1 Selection and sampling

The geographical location of where the research took place is considered a melting pot for autonomous driving technology. As such, we were able to attend a seminar as well as an innovation Bazaar (cf. Nambisan & Sawhney 2007) in order to get initial insights of the area as well as establishing contact with people that could be of interest to interview. In order to find more respondents relevant to the study, we also browsed through attending lists of previously arranged seminars on autonomous driving. We specifically set out to find respondents that were from different domains as that would provide a multidimensional view of the problem domain (Myers & Newman 2007; Patel & Davidson 2011).

Additionally, we also used the snowballing technique (Biernacki & Waldorf 1981; Widerberg 2002) to find respondents. We asked our respondents if they knew people from different domains that were of relevance for our study. The snowballing technique was used because the efforts and projects regarding autonomous cars at this location encompass people in joint efforts from a wide array of domains and fields. Most of the people involved know each other or are aware of the work of other individuals as well as their respective field of expertise. The traffic specialist, the mechatronics engineer and the signaling engineer were found this way. We thus found respondents that were involved in the development of autonomous cars, either direct through engineering or indirect through policy work or research. An overview of the respondents, their role in autonomous cars and how the empirical data was collected is portrayed in table 1 below:

In text	Role in autonomous cars	Collected empirical data
Politician	Policy work in autonomous car governance	Semi-structured, 30 minutes, recorded
Security specialist	Researcher, security aspects of autonomous car systems	Semi-structured, 30 minutes, recorded
Mechatronics engineer	Researcher in cooperative driving between autonomous cars and others	Semi-structured, 30 minutes, recorded
Traffic specialist	Traffic governance in public sector, involved in autonomous car projects on international levels	Semi-structured, 40 minutes, recorded
Signaling engineer	Researcher vehicular communication, signal processing	Semi-structured, 45 minutes, recorded
HRI researchers 1 and 2	Researchers HRI, specialists in autonomous car interaction	Semi-structured, 60 minutes, recorded
CarCorp executive	Governance and policy work	Semi-structured, 60 minutes, recorded

Table 1: An overview of the respondents that participated in the study

3.2 Data collection

As the locus of study can be very theoretical and confusing for those not familiar with agency as a concept, we had to find a way in which we were to get our respondents to discuss the concept without us mentioning it directly. As such, we figured scenarios would help us in this regard as they provide a visual and conceptual tool for indirectly discussing agency. We thus created four traffic scenarios using a modelling program online that provides free tools for illustrating traffic accidents (Accidentsketch 2017). This online tool is mainly used to illustrate traffic accidents for insurance claims but we found it useful for illustrating scenarios that pinpoints agential issues that arise when autonomous cars are introduced in traffic. The scenarios are displayed and explained in more detail below.

The scenarios were used as a form of semi-structured interviews (Bryman 2012) as each scenario can be seen as a question that can be explored. The four traffic scenarios highlights problems that humans effortlessly deal with as we are able to understand the intent of others, showing intent and cooperating socially. It is dubious, however, whether autonomous cars can deal with these real world traffic situations at the same levels of humans. We are moving towards a state when it may be possible to simulate the actions a human driver could do. And as such, is it advisable for autonomous cars to simulate human behaviour, or are the scenarios to be "solved" in different ways? The scenarios thus deals with different types of interactions that can arguably be solved in several ways. The respondent's solutions provide an insight into how agency is attributed to autonomous cars in the social environment of traffic. For instance, how much control that is given to an autonomous car in a specific situation would classify it having some form of agency.

In all scenarios, an autonomous car is illustrated in red while blue and beige cars houses human drivers or are parked. The first scenario deals with the interaction between an autonomous car and a human driver. The second scenario deals with the interaction between an autonomous car and pedestrians. The third scenario comprises interaction between an autonomous car and a human police officer, it also shows a situation that arguably is very unique in that it most likely will be different from one situation to the next. It could thus also be seen as a temporary road construction site. The fourth and last scenario combines both pedestrian and human driver interaction and is along with scenario three arguably most complex.

The figure text and the description of each scenario were not shown to the respondents in order to not steer their answers in any direction. What was shown were the different figures and its corresponding text in italics. Worth noting again is that the autonomous car is classified as being level 4 or 5 on the SAE J3016 standard (2016). That is, the autonomous car does not have direct human fallback.

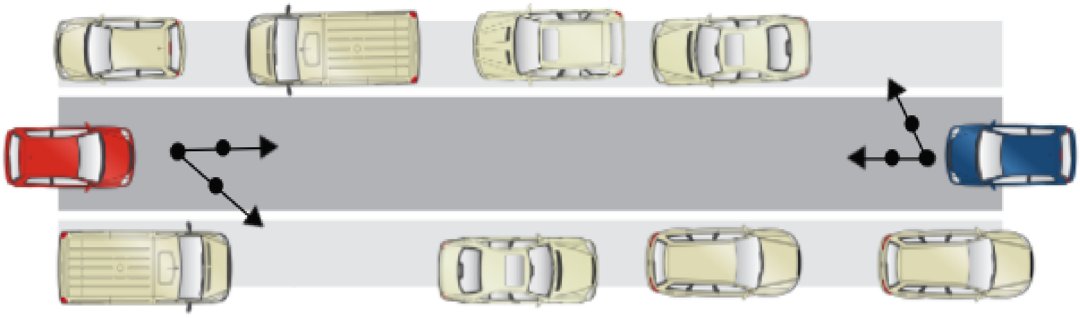

Figure 1: Autonomous car (red) and human driver (blue) interaction

Scenario 1: **Autonomous car and human driver interaction**
An autonomous car is approaching a narrow street with parked cars on both sides, there is only room for one car. The road has two-way traffic and there is a human driver approaching from the other side. There is only one car that can use the narrow passage at a time.

The first scenario concerns the interaction between an autonomous car and a human driver. Cooperation between the autonomous car and the human driver is usually needed to solve this situation.

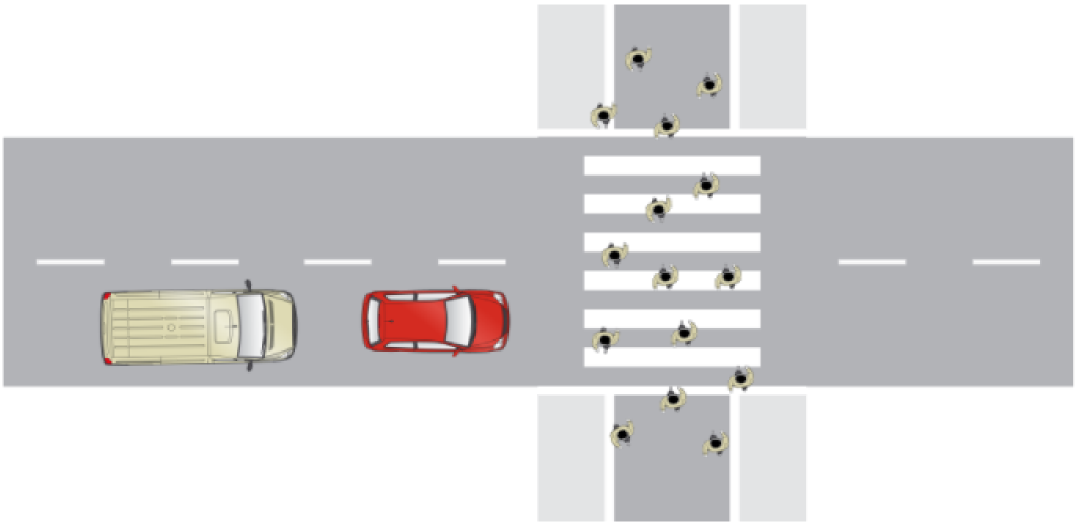

Figure 2: Autonomous car (red) at the intersection of a busy walking street

Scenario 2: **Autonomous car and pedestrian interaction**
An autonomous car is approaching a busy walking street. There is a constant flow of pedestrians crossing and other cars are approaching from behind and are getting eager to drive on.

The second scenario usually requires an understanding of the intent of others as well as having the ability to show its own intent.

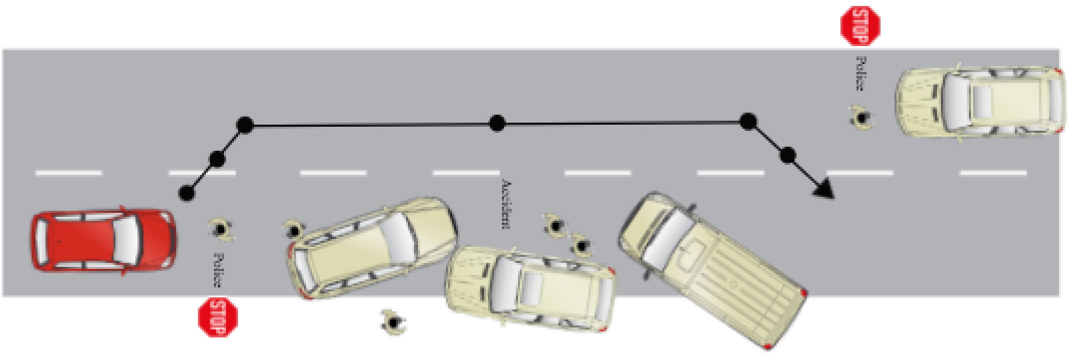

Figure 3: Autonomous car (red) and human (police officer) interaction

Scenario 3: Autonomous car and police officer interaction, unique situations
An autonomous car is approaching a situation where an accident has happened. A police officer by the stop sign is steering the traffic so that cars have to enter the opposing lane for the part of the road that houses the accident. Another police officer has stopped the traffic on the opposing side.

The third scenario requires an understanding of the signals given by the police officer and puts more to the point of how an autonomous car can understand intentions of humans that are not inside other vehicles. The scenario also pinpoints the uniqueness of every situation as these situations tend to vary, it puts questions regarding how much an autonomous car can actually understand of a situation that is close to impossible to foresee beforehand. For this reason, this scenario was also described as a road construction site when displaying it to the respondents to stress how a situation can be very ad hoc in nature.

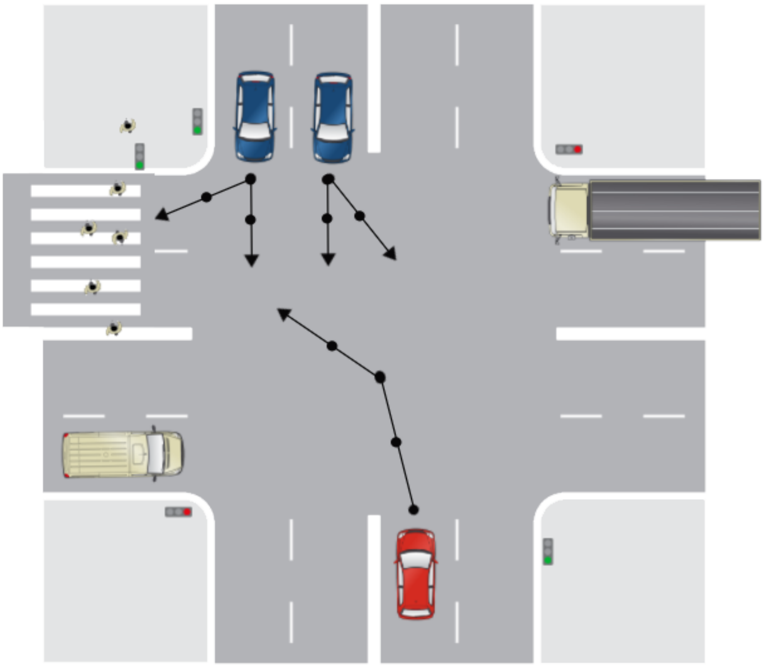

Figure 4: Autonomous car (red), human drivers (blue) and human (pedestrian) interaction

Scenario 4: Autonomous car interacting with both human drivers and pedestrians
An autonomous car is approaching a road crossing where it has a green light and is about to turn left. Opposing traffic (blue cars) also have a green light and might be turning left, driving straight ahead or turning right. There is also a zebra crossing where pedestrians have a green light on the left hand side.

The fourth scenario comprises a complex situation where an autonomous car is to interact with both human drivers and pedestrians. In this situation, an autonomous car has to understand the intent of others as well as showing intent on its own. Furthermore, it requires a sense of flow since the car cannot be stuck in the middle of the road.

In all of these scenarios, cooperation is required to lesser or more degrees. Intentionality is also required and the scenarios provide questions such as if an autonomous car should simulate this or solve the problem in another way.

Inspiration from the scenarios were found from the author's real world experiences of driving in traffic where interaction has been key in solving the problems at hand. We placed an artificial driver where a human usually would be. As such, we invoked questions whether an autonomous car itself should handle the situation, and if so, how? Moreover, maybe the car should not handle the situation at all, and if so, how could the situation be handled differently? To our best effort, we thus made the scenarios pinpoint interactions that would force respondents to address how agency is attributed to technology in different scenarios one would likely find in the real world.

Before the recorded interviews were about to take place, we e-mailed the scenarios to each respondent and shortly briefed them on what the interview entailed. Furthermore, we told them that their answers would be portrayed anonymously as well as confidentially in the study (Andersson 2001). We also asked the respondents if we were allowed to record the conversation, they all approved, and we thus used a cell phone as a recorder. Field notes were taken during the interviews and were used as a complement to the recordings at a later stage. A total of seven interviews were conducted with eight respondents where each interview lasted 30-60 minutes. Each interview were held in Swedish or English at the respondent's workplace. One interview had two respondents present where one of them was available via Skype, their answers are treated separately.

The scenarios were sequentially shown to the respondents as we asked how the respondents would solve the scenarios, not merely asking how an autonomous car would handle the situations as that would limit the answers to technical solutions. The questions we asked during all interviews revolved around the same basic topics (see Appendix 3). Moreover, we touched upon some general issues in each interview. By doing this, we could get some insight into how the respondents viewed autonomous driving in more general terms, such as adoption rate and viewing traffic as a social environment. These issues were not the main locus of our study, but helped us establishing trust (Myers & Newman 2007), validating our theories and gaining valuable insight into how the concept is approached and viewed from each respondent.

Using the scenarios, we were able to explore the theoretical assumptions regarding how agency was attributed to autonomous cars. To exemplify: if a respondent would propose that the autonomous car should seek help from a human in a scenario, we could classify it as having technological agency as found in IS research, where it is considered a tool that humans use. And contrariwise, if a respondent was confident that the autonomous car should solve a situation all by itself, it could indicate that the car itself had agency that is usually ascribed to humans or AI. Furthermore, using the scenarios also examined the practicalities of adding autonomous cars to the social environment of traffic.

The traffic scenarios were validated through use as many respondents found the different situations likely to occur in the real world as well as the relevance of the difficulties they propose. To our best effort, we let the respondents speak freely about how to solve the scenarios without us interfering or giving aid in how others had solved the issues in order to not taint the results (Andersson 2001). This was done while still keeping them close to the topics seen in Appendix 3. As such, it was possible for us to identify where the respondents had opposing views to one another as well as identify patterns and themes since all the interviews were based around the same principle.

3.3 Data analysis

Since our aim was to understand patterns between the different stakeholders in our empirical data, we found the thematic analysis appropriate (Braun & Clarke 2006). We began the data analysis process by transcribing all seven recorded interviews. This resulted in 61 pages of text. As we read them through, we compared them to the field notes taken to ensure that we did not overlook anything of importance. To reduce overlap, we compared different concepts to each other and by doing this, we were able to find some redundancy. For instance, concepts such as "control center", "control tower", and "third party controller" were merged together.

We analyzed the empirical data in an iterative process and began reading through and coding text sections into themes. The themes used were based on both theory and what was found in the empirical data (Widerberg 2002). As such, the themes highlighted different aspects of our theoretical assumptions (Bryman 2012). This work involved printing out all transcribed interviews where we cut text sections as codes with a pair of scissors and then sorted relevant codes together. The codes were then re-read several times where we grouped them to other similar themes and codes while constantly evaluating them for relevance to the scope of the study (Bryman 2012). An overview of the themes can be seen in Appendix 1.

As the themes were established and worked through, we also noted relevant ideas and connections between the theories presented (Braun & Clarke 2006). At times, we found theoretical implications that we had not touched upon in the study previously but were still deemed relevant to the study. For instance, all respondents mentioned infrastructure in relation to autonomous cars and we thus found it important to include this in the study. The themes were evaluated a second time and condensed from the initial 14 found in Appendix 1 to the six found in the empirical findings section below. Details of which themes that were merged can be found in Appendix 2. The empirical findings were then written in natural text and representative quotes were used interspersedly. We translated the quotes to English when necessary. The results are found in the empirical findings section below.

4. Empirical findings

The results are presented in line with the themes found during the data analysis (Appendix 2). Initially though, we present some general results that were found during the interviews before discussing the actual scenarios.

4.1 General observations

All of the respondents thought that there will be a slow transition towards autonomous cars and that mixed traffic will occur. Moreover, all respondents identified traffic as a social environment where different actors must interact for it to function properly. Furthermore, safety was a prioritized concern according to the HRI researcher 1, the CarCorp executive and the politician where the “human factor” in accidents can supposedly be reduced with autonomous cars. Lastly, the actual value of autonomous cars were not that easy to specify for either of the respondents.

The first scenario invoked discussion how the autonomous car could interact with the human driver in order to proceed. Most respondents agreed that the interaction would be easy if both cars were connected digitally through vehicle to vehicle (V2V) communication. However, the CarCorp executive deemed this a bad approach as you can never be sure that every other car uses the same or any V2V technology, reasons being a slow adoption rate of autonomous cars and that it will be in a mixed environment with non-autonomous cars. Market penetration of V2V were mentioned as a problem by the mechatronics engineer, the signaling engineer and the security specialist as well. The mechatronics engineer stated that:

> *“The problem is when one of these two vehicles is human and we don't have any [V2V] communication.” - Mechatronics engineer*

The respondents thus agreed on that the autonomous car would have to understand the intent of the human driver in some other way.

4.2 Designing passive or aggressive behaviour

The CarCorp executive was hesitant if the autonomous car should handle situations like the first scenario at all, as the ability for machines in understanding human intentions and enacting complex coordinated movement is limited. The respondent stated that humans would solve this situation quite easily. The signaling engineer and the HRI researcher 1 argued that the situation could be solved by the car using its body language. The signaling engineer described that:

> *“I presume they [the autonomous car and the human driver] don’t communicate with each other [with V2V], then you have to use body language.” - Signaling engineer*

HRI researcher 1 & 2 claimed that in these situations, one car often must make the first move, and one could thus program "aggressive" behaviour to make the first move:

> *"Well, someone has to make the first move [...] if not, they will stand there for an eternity." - HRI researcher 1*

> *"At the beginning, these cars were having difficulties driving autonomously but as they drove more aggressively, they claimed more space and thus took the initiative to drive first." - HRI researcher 2*

The signaling engineer proposed that probing technique could solve the situation, where the autonomous car would move forwards and see if it gets a reaction from the other car in form of movement, this process would be iterative where options and responses are constantly evaluated.

In the second scenario, the signaling engineer and the HRI researcher 1 similarly argued, as in the previous scenario, that the autonomous car could probe the pedestrians and see if it gets a reaction to pass. The former stated that human drivers have this behaviour and that autonomous cars could adopt it and stressed though that the car must move very slowly to not harm people and that any sign of people not stopping should prompt the autonomous car to halt. However, the respondent thus mentioned that we do not know whether society will accept this behaviour. The traffic specialist was hesitant to this approach as one would not want to program challenging behaviour into autonomous cars. The politician shared this point of view and thought a reasonable approach for the car would be to wait to get a clear path:

> *"To work in this system, the [autonomous] car has to be humble." - Politician*

The CarCorp executive was also extremely hesitant to the approach of probing as the autonomous car must act politely and passively in all situations. One of the reasons being that there are laws stating that pedestrians have the right to be on the crossing and programming this behaviour would be legally wrong:

> *"Current laws state that pedestrians that have set foot on the crossing have the right to be there. [...] The [autonomous] car is to be careful, polite and always obeying the law." - CarCorp executive*

The CarCorp executive further referenced to another firm that had tried similar probing behaviour only to receive massive public outrage as a result. This, of course, could have implications the respondent argued. For instance, an autonomous car does not get stressed if it has to wait longer, which might be annoying for other road users if it acts very passively and hinders traffic flow.

4.3 Indicating if cars are autonomous

As the autonomous car deals with human interaction in the second scenario, a topic that was addressed was whether the autonomous car should in some way signal that it is in fact autonomous. Whether or not this should be the case was being discussed in the field according to all respondents, where no real consensus has been reached. The politician argued that you should tell people it is, because it might be ethically wrong to not do so since they might function on different criteria than human drivers. The HRI researcher 2 and the security specialist, however, both highlighted possible problems in showing it is autonomous since it might create over-dependencies where pedestrians might trust too much in the technology and thus exploit the autonomous cars. For instance, if they know that autonomous cars will always stop, they might just walk out in front of it, the HRI researcher 2 exemplifies with the thought process of a pedestrian:

> *"It's automatic, so it should stop when i walk out in front of it." - HRI researcher 2*

The respondent flagged that this might be a problem if the car's sensors might not work, for instance, or if other cars behind would make it impossible for the autonomous car to stop. The politician argued in a similar manner that it is troublesome if pedestrians challenge cars since the strength balance is heavily favoring the car:

> *"A pedestrian cannot challenge a car, it's like comparing David and Goliath."*
> *- Politician*

The CarCorp executive, however, did not see this as a problem since it is very unlikely that people would take the decision to walk or jump out in front of an approaching car. The respondent highlighted that as this is deeply enrooted in our human minds, we would never take such chances:

> *"To jump out in front of a two-ton car [...], it's probably easy to say, but to do so in practice…" - CarCorp executive*

According to the respondent, the Japanese government lifted the ban of autonomous cars in regular traffic for this very reason.

Between other cars, there are issues of challenging the autonomous cars as well. The politician and the CarCorp executive referred to the same event where other road users had challenged an autonomous car in trying to force it off the road with their own car. The latter explains:

> *"Other road users who saw that they [CarCorp's autonomous test cars] were autonomous drove in front of them to make them brake abruptly. There were also cases when they [other road users] tried to push them off the road."*
> *- CarCorp executive*

The respondent speculated that these situations might have occurred since the technology is quite new and that this behaviour will most likely diminish once the novelty of autonomous cars in regular traffic has worn off.

4.4 Understanding humans, pedestrians and other vehicles

Another topic that was raised was how the autonomous car can understand human behaviour. The signaling engineer argued that the autonomous car is occupying the same reality as us humans, and as such, it is exposed to the signals and all information that humans can perceive, and it is therefore up to technology to interpret and predict human behaviour.

> *“[...] the information is right there. It’s there since we can interpret it and therefore the autonomous car should be able to do so too.” - Signaling engineer*

The mechatronics engineer had a slightly different view on the subject since human behaviour cannot be predicted by mathematical models. This however, would only apply to pedestrians as they are very unpredictable and can move in any direction at any given time. Vehicles, on the other hand, are constricted in the way they can move. They are restricted physically as they cannot accelerate or stop ingeniously fast. As such, it is not of interest to understand how the person behind the wheel is planning how to proceed, since the possible states of the vehicle can be predicted by mathematical models. Pedestrians, on the other hand, can move in every axis of movement, making their movements harder to predict according to the respondent:

> *“That's problem with the humans. You don't have a mathematical model to predict them while with a vehicle, okay, you have pretty good mathematics describing what is the motion.” - Mechatronics engineer*

Moreover, the mechatronics engineer claimed that it is easier to predict human behaviour if they are at a zebra crossing as opposed to just standing at the side of the road or walking on a sidewalk, since the zebra crossing would signal that they are actually going to cross. These are issues that can be enhanced by the use of reinforcement learning in AI. For instance, the autonomous car could use the information that people at zebra crossings tend to cross them and connect that knowledge to third party information saying that there are many pedestrians at its current location. However, there are problems with AI according to the mechatronics engineer since you cannot predict the outcome of an AI algorithm:

> *“[...] the problem with machine learning, with AI, is that you don't have any way of understanding the decision, the output of your AI-device.” - Mechatronics engineer*

For this reason, they constrict the use of AI to predictions, not to actual decisions in handling the autonomous car. AI is thus used to other things that can not be calculated by mathematical models, such as human behaviour. It is, however, still a challenge in understanding pedestrians with AI according to the mechatronics engineer due to them being highly unpredictable.

Furthermore, in understanding human behaviour, the CarCorp executive, the HRI researcher 1 and the security specialist similarly noted that driving behaviour differs between geographical locations. The security specialist exemplifies that even local variations of driving can be observed in the same country, such as giving priority to the right is different between city traffic and countryside traffic:

> *"The right hand rule works somewhat, if you stick to it [in Gothenburg] you get run over. However, in other locations, such as the countryside and other cities it's the opposite, if you don't give way, you'll hit the side of someone who thinks - 'it's my right to drive here'." - Security specialist*

In a similar way, the HRI researcher 1 exemplified that just because an autonomous car has learned how to drive in California it does not automatically mean that it can drive in Gothenburg.

4.5 Third party control

In the third scenario, the autonomous car is approaching a very complex situation and the respondents came up with different and creative ways in handling it. The CarCorp executive, the traffic specialist and the HRI researcher 1 all thought that the car should stop and ask a third party for help when it came to a situation it could not understand. The main idea is that a third party traffic controller can overtake the control of the car if needed as the traffic specialist mentioned:

> *"It's very possible to imagine a future where a third party participates and takes control over certain situations." - Traffic specialist*

This was described similarly to airport traffic control towers. It was also a matter of what autonomous level the car is on, if the car is on level 4, the car can stop and ask the human in the car to regain control in a structured way according to both the security specialist and the CarCorp executive. When there is no human present, that is at level 5, the third party traffic controller method was advisable:

> *"It [the autonomous car] can't do anything here [the third scenario], it can only wait them out. It could send a signal to a control tower [...] to require assistance from someone." - CarCorp executive*

The HRI researcher 1, the signaling engineer and the traffic specialist all stated that a possible solution to the scenario was to equip the police officer with a device that was connected to the vehicle digitally and gave the autonomous car a new trajectory. The signaling engineer and the traffic specialist exemplifies:

> *"One could equip the police officer with a small communication device that is used to give instructions [to autonomous cars]." - Signaling engineer*

"[...] One could imagine that a police officer's signals could be amplified by some type of digital transmission [...] that could send additional information [to the autonomous car] than just the purely visual one." - Traffic specialist

The HRI researcher 1 further stated that the amounts of possible hand signals and body language of a police officer are limited and that autonomous cars could be able to distinguish these in the future:

"It's possible to make systems [autonomous cars] understand police officers' hand signals since the actual amount of different hand signals are limited." - HRI researcher 1

The CarCorp executive identified problems when an autonomous car should understand hand signals, as the autonomous car might just think that it is a pedestrian that it should avoid. In this way, it could just drive past the police officer instead of stopping as it is supposed to:

"And a self driving car says – Oops, let's not hit this pedestrian [about the police officer], I'll drive around and then continue driving on." - CarCorp executive

As the third scenario also could also be described as a construction site, the signaling engineer thought the road construction workers could place a digital beacon telling the autonomous cars to take a different path, however noting that it was different in terms of being planned or not:

"If this was to be a construction site instead, a planned activity, then you could place a communications device that describes a path past the construction site." - Signaling engineer

The politician identified vast problems since these road construction sites are often ad hoc in nature where the workers often forget or do not even bother establishing road signs indicating that road construction work is underway. Furthermore, the city had no current method in digitally signaling that road construction work is underway:

"Today, we have troubles with some of the contractors not using signs adequately to show that road construction work is underway." - Politician

The respondent, however, had faith in that as autonomous cars rely on real time maps technology, they could avoid these situations as maps could show these occurrences in the future.

4.6 Altering the infrastructure

The fourth scenario was identified as easy to solve by some and less easy by others. The CarCorp executive thought the autonomous car should similarly to the previous scenarios wait to get a clear path before driving on. For instance, the traffic specialist deemed it simple and thought it was down to regulating the signal posts. However, one should not rely solely on signals according to the respondent, as humans can run red lights or jaywalk. This was similarly identified by the signaling engineer who argued that the autonomous car must take this into consideration and thus be careful:

> *"It doesn't matter if pedestrians have a red or a green light because you never run over a pedestrian." - Signaling engineer*

The way car manufacturers are dealing with situations similar to scenario 4 is to avoid them completely according to the CarCorp executive. One of the reasons being that there are too many parameters that are out of their control in these situations, such as people running red lights or humans that acts unpredictably:

> *"Then we cut off [where the autonomous cars are in control] at intersections and roundabouts. Our roads are then: Clear lane markings, no intersections, no pedestrians or cyclists." - CarCorp executive*

The approach of limiting autonomous cars to certain traffic situations is something that was mentioned by other respondents as well. The traffic specialist claimed that one could avoid the problems that arose in the scenario by separating the cars and the pedestrians:

> *"First and foremost, I believe it's up to separating the rule based and the non-rule based acting in time and space." - Traffic specialist*

Moreover, the CarCorp executive thought this could be managed on a policy level, where cars were separated from other forms of traffic such as pedestrians and cyclist based on the allowed speed:

> *"Above 40km/h you could separate everything and you must either regulate or separate it by level differences."* - CarCorp executive

Similarly, the politician saw that in the distant future, one could assign certain parts of the city for autonomous cars and vice versa, in some parts they would not be allowed at all. The security specialist thought that autonomous cars could be given assigned lanes on freeways, where no regular cars are allowed:

> *"It's not unlikely that some roads such as highways will only permit autonomous vehicles in certain lanes, much like bus-lanes or the like." - Security specialist*

The signaling engineer saw a possible solution in better adapting the infrastructure to autonomous cars:

> *"One can look at what can be done to the infrastructure in order to facilitate for an autonomous car." - Signaling engineer*

The politician, however, stated that the infrastructure will not be altered to fit autonomous cars, instead autonomous cars have to be adopted to the infrastructure:

> *"We neither can afford nor have the opportunity to rebuild the city [infrastructure] to support autonomous cars. Instead, the autonomous cars have to adapt to the city." - Politician*

The traffic specialist and the CarCorp executive claimed similarly that adapting the infrastructure to technology is not something that would work on a global scale as most of the infrastructure is already built. The traffic specialist claimed:

> *"I don't think a universal solution would be to adapt the infrastructure to technology as that isn't particularly sustainable on a global scale. [...] most of the infrastructure is already built and won't be altered [...] so developing technology that relies on rebuilding the infrastructure will be difficult to apply." - Traffic specialist*

5. Discussion

This section is structured as following: The themes from the empirical findings are discussed in relation to the research question with an overall structure that reflects the gathered data. As the scope of the study is twofold, the discussion section first addresses autonomous cars in the social environment of traffic followed by what these findings say about how agency can be attributed to autonomous cars. This is portrayed in figure 5 below. The results are discussed in relation to the theory presented together with other relevant theories and research that can help in understanding and explaining our findings and answering our research question. Finally, we state contributions to theory and practice, limitations of the study and areas for further research.

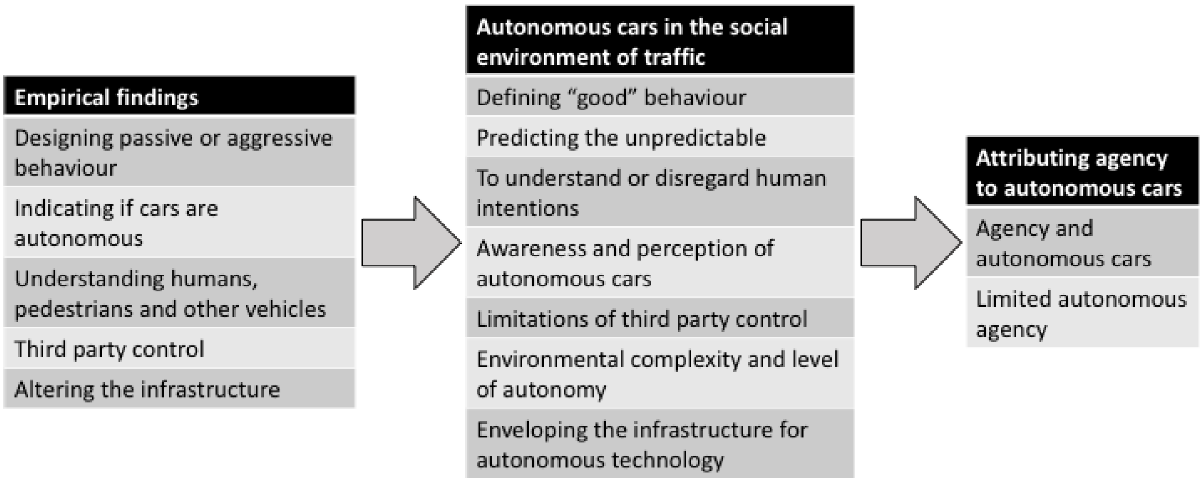

Figure 5: Layout of the discussion chapter

5.1 Defining "good" behaviour

A common argument one often comes across in the area of autonomous cars is that humans account for most accidents and subsequently removing the human component should lower the overall accident rate (Kröger 2016; Näringsdepartementet 2016). While this might be true, we will still most likely not find ourselves in a situation where all cars are autonomous (Fraedrich et al. 2015), rendering the argument less applicable. This puts more relevance towards the scope of this study in that the autonomous cars must find alternate ways to communicate with and understand humans.

Some respondents from academia proposed that the autonomous cars could use probing behaviour where the cars were given a certain amount of freedom to show with its body language what it was about to do in both scenarios of interaction between regular cars and pedestrians. While probing, it would also evaluate its actions, if it gets a response. Further, they thought that you could have the cars mimic how human drivers behave to achieve a natural way of driving and thus a natural way to behave in mixed traffic. To succeed in this, they stated that the autonomous cars should only learn the "good" behaviour of humans. We would argue that it is easier said than done. For example, in considering the roads as social environments, formal rules are often broken or bent on a regular basis, contingent to the situation at hand. In some instances, some rules can be broken, whilst in others, the same rule cannot. We thus find it problematic to define what is considered "good" behaviour in this sense at it differs widely from different situations and that in the real world, we are put in front of situations that are infinitely unique. Further, difficulties in having a computer distinguishing between "good" and "bad" behaviour is nothing new. For instance, IBM's supercomputer Watson had trouble in understanding the difference between profanity and polite language after picking up bad habits from Urban Dictionary and Wikipedia (cf. Lev-Ram 2013).

Moreover, in taking geography in respect, many respondents identified that there are local variances in driving. It is not hard to imagine that an autonomous car that does not take this into consideration would not to be regarded as a functioning actor in its environment (Juhlin 2010). But how would it do this? Once again, we enter the discussion of what is considered "good" behaviour. Should we program a car to be able to drive in suburban or metropolitan areas? A possible solution could be that the autonomous car alters its behaviour depending on its geographical location. The difficulties though would arguably be to define what constitutes the behaviour of a certain location. It might be hard to program that "sometimes, on the countryside, people do not obey the right hand rule" as identified in our empirical data. Moreover, it is not unlikely to think that cars with a different origin are visiting another geographic location. In this case, one could have the autonomous car mimic the local variance once it is there. But the problem still persists that not all other cars will in fact behave in a predictable manner even though most cars will generally behave a certain way. This prompts the autonomous cars to have an overall very careful approach as it can never be sure of how any other actor will behave, possibly becoming a nuisance to other drivers which we will return to later on.

5.2 Predicting the unpredictable

The discussion regarding what is considered "good" or "bad" behaviour can be extended to how autonomous cars could understand humans on a general level. In this study, this was most prevalent in the third scenario where an autonomous car were to understand the instructions of a police officer or a flag guard. The HRI researcher 1 argued that there is a limited set of hand signals a police officer can use and that there is not really a problem in getting an autonomous car to understand them. This is an interesting idea, but one that we would be hesitant to agree upon. First off, a police officer can be visibly very different between individuals, perhaps there is a police officer off duty that has to direct traffic, or maybe the police force just changed their uniforms. Perhaps police officers have yet to arrive and people from the fire brigade have to direct traffic in the meantime. Second, hand signals and body language do vary widely between individuals and countries. For instance, if someone in traffic in central and northern Europe as well as in the US nods his or her head, it usually means yes and shaking one's head usually means no. There are exceptions, however, whereas in countries such as Bulgaria and India, shaking one's head sideways, a motion akin to the meaning of no as just described in most European countries and the US, actually is a gesture that signifies agreement (cf. Färber 2016).

Moreover, as identified in our empirical data, humans cannot really be predicted by mathematical models accurately that further problematized formalizing a standard set of hand signals. As discussed before, you cannot really predict how all humans will behave in a certain way on the basis of what most people do. This is evident in other research as well. In experiments with machine learning at MIT, machines scored significantly lower than humans in predicting the outcome of human interactions, proving that both humans and machines have trouble predicting human behaviour. Machines, however had more trouble doing so as they seemingly lack the subtle skills involved in judging complex human interactions (Conner-Simons & Gordon 2016).

5.3 To understand or disregard human intentions

We could thus ask if it is appropriate to give mandate for autonomous cars to simulate social agency (Leonardi 2012; Pickering 2001) by using probing behaviour as proposed by some respondents from academia. This provides ethical questions as we can argue that technology is different than humans in the sense that it does not have qualities attributed to humans such as intuition or courtesy (Floridi 2014; Rose et al. 2005). As such, they would only simulate that they understand intentionality. They could mimic humans, but they would only do so based on programming, not judgement. As such, they would not truly understand situations, again proving difficulties as we expect them to have good judgement in taking adequate decisions in complex situations involving human beings.

The mechatronics engineer had an interesting view regarding this subject where one should not need to understand human intentionality when dealing with traffic as one should just be interested in predicting how objects move. As cars are limited by physics, they cannot stop at an instant or make turns that are physically impossible given the current state of the vehicle. In understanding traffic like this, things becomes a lot easier as one only have to deal with objects and their proposed trajectory. However, if we perceive the traffic environment of today as highly dependent on cooperation (cf. Brown & Laurier 2017; Juhlin 2010) it could be troublesome if several actors in this environment did not aim to cooperate with others, only following its own path and avoiding moving objects. We thus ask the question if it is advisable to disregard understanding intentionality in an environment that is reliant on interactions to function. The observations from Brown and Laurier (2017) pinpoints problems that arise with this exact issue, when autonomous cars do not understand or show intentionality properly, prompting us to be hesitant that this is a viable approach.

As previously discussed, the concept of flow is reliant on cooperation (Juhlin 2010) and if taking a wholly passive approach one would get autonomous cars that hinders flow. In the other extreme, if they are aggressive where they simulate that they understand intentionality and acts like humans would, we once again touch upon the problems in defining what is considered "good" behaviour. We thus find ourselves in a situation where we need to balance a passive to an aggressive approach. As some respondents said, sometimes it is advisable for an autonomous car to be aggressive while in other situations it might not. Humans use intuition and judgement in every situation and the question is whether one could assign this responsibility to the technology as it inherently lacks the traits associated with making these decisions (Floridi 2014; Rose et al. 2005). Research have been done extensively by Google and programming cars to be aggressive or "assertive" tends to make them function better in traffic (Gray 2014). It is however dubious, if it is viable of programming a flat rate assertive behaviour this way, as research made by Google and subsequently the firm Waymo is at the moment not released publicly.

5.4 Awareness and perception of autonomous cars

With the introduction of autonomous cars in traffic, we find ourselves dealing with perceptions of the autonomous car and whether one should be able to tell if it is autonomous or not. This is important if we take challenging behaviour into consideration. As identified in our empirical data, there are currently no consensus regarding if one should be able to tell it is autonomous or not. Academia and practice had in this regard contradicting ways of addressing the issue. Academia claimed that people could take advantage of the autonomous cars and create over-dependencies, such as knowing that it will always stop if one walks out in front of it. Practitioners thought, however, that people would never do such a thing as it is not intuitional. We would however argue that there are certain situations that one could take advantage of the autonomous cars. For instance, if the car is already standing still and give no clue as to moving forwards as proposed in most scenarios by the CarCorp executive, the majority of people would probably walk out in front of it. After all, if there is no driver sitting in the car, who cares if it gets to wait a little bit more? The politician stressed that it might be troublesome as people do not know what criteria an autonomous car is driving under. It could again be the problem we will go into more detail later, regarding the inability of stating the overall goal of the vehicle.

Furthermore, there is also the danger of other drivers challenging the autonomous cars, but arguably less so, as both parties are vehicles. It seems to us that people are curious in how autonomous cars functions, and as such, they want to test its limits. Other road users do not know what premises they run on. The autonomous cars might come off as strange as they are given the same dignity as humans and as such, it is tempting to see how it reacts in different situations. It could be a natural way for people to understand how it acts in "their" environment. It is after all an alien technology that enters the domain of human drivers, so the people in it seemingly seek to understand what premises it functions under. As such, they test it by action. We would argue that it is not unlikely that pedestrians would do this as well, but once the premises are clear, this behaviour would probably lessen. This would also require that the cars do function under clear premises, something that today is rather unclear.

To understand this interplay better, one could delve into the field of anthropomorphism as it deals with the "*tendency to attribute human characteristics to inanimate objects, animals and others with a view to helping us rationalise their actions*" (Duffy 2003, p. 180). The author argues that robots in social settings should only exploit the expectations of the robot in having human characteristics than actually strive to be perceived as human-like. As such, anthropomorphism could help in fine tuning the interaction between human and technology. However, as identified in our empirical data, making technology resemble humans does not come without problems as it might create false expectations if it comes off as human while not sharing the characteristics of humans (ibid). We will not go into this debate any further, but leave this a subject for further research.

5.5 Limitations of third party control

If the autonomous car approaches a complex situation it cannot solve by itself, our empirical data suggested that it could ask for help from a third party traffic controller. However, this might be troublesome as it would probably take some time before a third party traffic controller could understand the situation and make a viable decision, thus clogging up traffic. Another option was to give control over to the proposed human driver in an orderly manner, this is probably a better option as the person in the car probably can solve the situation faster. However, as there are to be autonomous cars without human drivers, this method is not always applicable.

An even better solution is arguably to avoid the situation in the first place as identified in our empirical data. As the cars are to share real time maps, this could probably be done. However, there are roads that you cannot detour from which entails that troublesome situations cannot be avoided at all times. Moreover, some researchers thought you could flag on the map certain road construction sites and so on, however, there were no current ways of doing this according to the politician and it is unlikely that it will function in the real world as they are struggling at the moment to even get construction workers to establish signs when they work. Likely, however, the real time maps could by showing real time road traffic data based on slower traffic, prompting the autonomous car to take another route if possible. This have been done by Google with great results for some time by harnessing information fusion through big data, providing users of their map service with real time congestion information (Stenovec 2015).

In the case of a police officer or flag guard stopping the autonomous car, some respondents stated that the police officer could be equipped with a device that was connected to the vehicle that could give the autonomous car a new route in order to proceed. We see that in some scenarios, such devices might work, but in reality, it would probably not be feasible. First of all, it would have to be a device that all police officers would have to use in the case of an accident that is very ad hoc in nature. Moreover, there would be significant costs to equip all police officers with one extra device, especially when the market penetration of autonomous cars will most likely be low (Fraedrich et al. 2015). Further it would have to be a standard that is incorporated in all autonomous vehicles, another cost that entails some sort of global standard that manufacturers are forced to use. While this might be possible, it is not plausible. In light of this, we deem the device-proposition not probable.

5.6 Environmental complexity and level of autonomy

The way the CarCorp executive were dealing with the scenarios presented were to avoid the situations entirely. That is, their autonomous cars are limited to a narrow set of traffic situations they can handle. At the moment it is only capable of operating on two-lane freeways without intersections or roundabouts. The reason being that these freeways offers a highly controlled environment where the risk of unforeseen events are arguably very low compared to say, city or even countryside driving. However, what is noteworthy about autonomous cars in general is that it is contrary to prior autonomous technologies to operate in environments that is highly contingent, open and social. That is, in environments with high complexity.

As presented by some respondents in our empirical data, a way of dealing with the impending interaction problems was to separate the autonomous cars from human interaction. For example, one could separate the roads so that only cars would interact with other cars. It would be easier for the autonomous cars to only have to take other cars into consideration as the trajectory of vehicles can be predicted contrary to that of pedestrians as previously discussed. This could be done by level separation of roads so that pedestrians were not allowed to occupy these places at all or to signal regulate so that autonomous cars only interacted with other vehicles. What could be said generally though is that all of these efforts revolves around lowering the environmental complexity of where the autonomous cars are to operate.

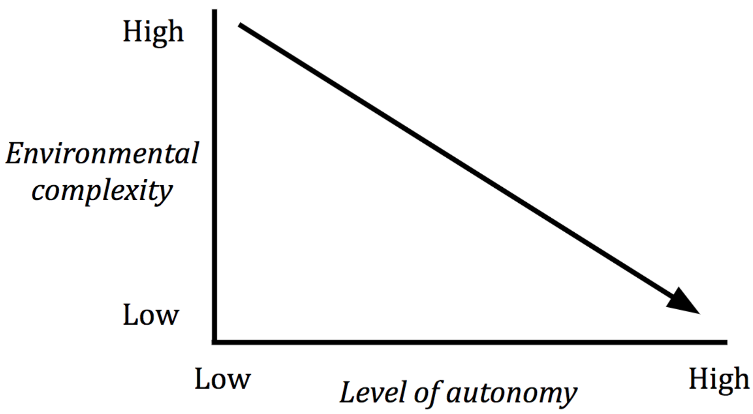

Figure 6: The interplay of environmental complexity and level of autonomy

The interplay of environmental complexity and level of autonomy can be described as figure 6 above. As the level of autonomy of a technology increases, the environment where it operates needs to be less complex in order for the technology to function autonomously. In our empirical data, we can see examples of this where the CarCorp executive explains how they strive to lower the complexity where the autonomous cars operate in limiting it to certain geographical locations that houses only a limited set of traffic situations. It is also reflected in the ideas of separating autonomous cars from other types of traffic or assigning certain lanes or parts of the city to autonomous cars. Moreover, we would find traffic as it is today to the far top left of figure 6 above, displaying high environmental complexity and a low level of technological autonomy.

5.7 Enveloping the infrastructure for autonomous cars

As such, it would seem that in order to make autonomous cars work, one must lower the complexity of the environment it is to be situated in. In the scope of autonomous cars, this would entail altering the infrastructure. This would classify as we are actually enveloping the environment for autonomous cars, but what does that mean?

Floridi (2014) claims that technology is not intelligent in a human sense, it is merely perceived as increasingly intelligent as we adapt our world to it. That is, we are setting up more rules so it can function properly in our environment. He exemplifies with autonomous lawn mowers, saying that they are as “stupid as your old refrigerator” (p. 136), but as we establish borders around areas for them to mow, they work rather flawlessly. In effect, they are labelled intelligent but what actually has happened is that their room to operate is narrowed down to the extent that they are perceived as being intelligent. The author calls this *enveloping*, when we adapt our environment to the technology. As identified in our empirical data, a similar approach has been applied when introducing autonomous cars in traffic. In order for them to function autonomously, they are limited to geographical locations with a low level of complexity that houses a narrow set of traffic situations they can deal with. The concept of enveloping can be described as figure 7 below:

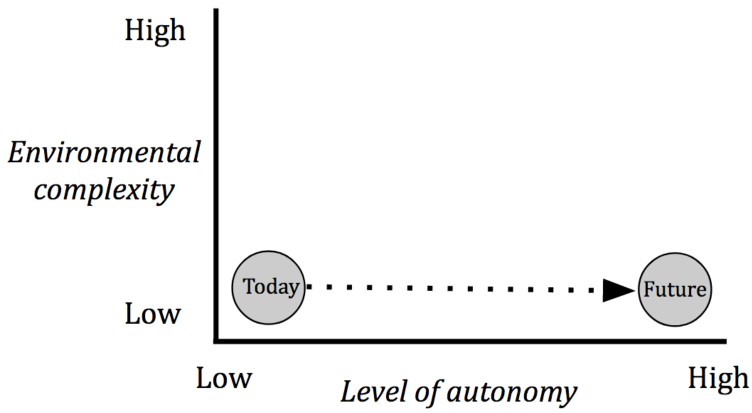

Figure 7: Current development of autonomous cars: Adapting the infrastructure to technology

Autonomous cars of today have low level of autonomy and are operating in environments with low complexity. That is, they are operating under very controlled conditions in environments with low level of unforeseen events that occur and few other actors to take into consideration. Moreover, they only deal with a narrow set of situations, arguably placing them as having a low level of autonomy while operating in environments with low complexity. The current trend is pointing towards further reducing environmental complexity in order to increase the level of autonomy as seen in our empirical data by separating the rule based from the non rule based.

The idea of altering the infrastructure to house autonomous cars, however, goes in stark contrast with what is actually feasible and likely to occur in the sense that infrastructures are not to be altered in order for technology to work. This is reflected in our empirical data directly by the politician and the traffic specialist. It could moreover be deemed as non probable considering that the value of autonomous cars are heavily debated, making infrastructural changes hard to justify on unclear grounds. Moreover, infrastructures are inherently hard to change (Ciborra & Hanseth 2001) which further complicates altering it to house a certain technology.

An idea found in the empirical data as well as in theory (cf. Anderson et al. 2016; Holmberg et al. 2016) was to assign certain areas of cities or lanes on freeways to autonomous cars. This could however prove to be quite provocative. To exemplify, autonomous cars are probably going to be quite expensive to either ride in or own. If the government was to favor certain areas for vehicles that only wealthier people could afford, there would probably be public outrage. We thus find ourselves in a contradiction where technology affords the infrastructure to be altered in order to function, but altering the infrastructure is something that is not feasible in practice.

The alternative to enveloping the infrastructure is then to make autonomous cars adapt to the social environment of traffic we find today. This can be illustrated as figure 8 below where autonomous cars are to adapt to and function in environments with high environmental complexity.

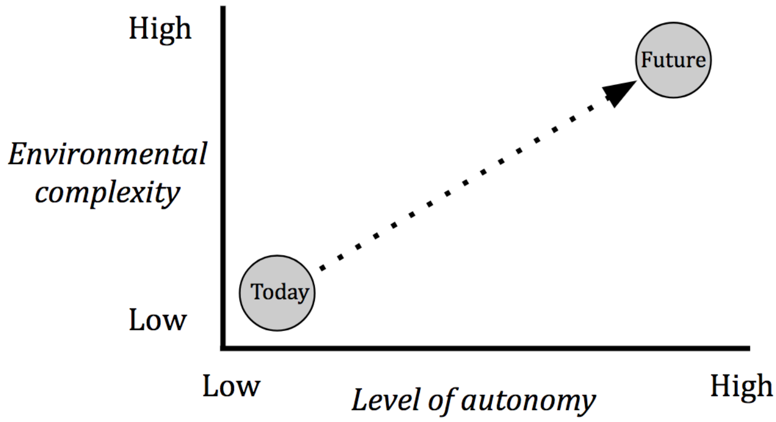

Figure 8: An alternate strategy: Adapting autonomous cars to the social environment of traffic.

If autonomous cars are to be fully autonomous in this environment, they must be able to understand and show intent and in effect enact social agency as previously discussed. More on this and in turn how agency can be attributed to autonomous cars is addressed in the following section.

5.8 Agency and autonomous cars

When we are to directly address how agency is attributed to autonomous cars, we can identify two main differences in the empirical data: The autonomous cars were in some instances given mandate by academia to show intentionality and understand the intentionality of humans, something that is often described as the main points of enacting human agency (Leonardi 2012; Rose et al. 2005). Further, the probing behaviour could arguably be an expression of social agency (Leonardi 2012; Pickering 2001) as the autonomous cars would to some extent cooperate with humans on own volition by probing to get a response. Contrariwise, we could distinguish that practitioners and respondents from public sector indicated that autonomous cars were to operate more passively, only on premises from those who created them. The laws that are being passed (cf. Kang 2016; Näringsdepartementet 2016) and statements from the firms themselves (Volvo 2015) reflect this notion where the firm takes full responsibility when the vehicle is in control of the driving task. In this case, it is very clear that the autonomous cars would have agency as described in classical IS-fashion. That is, as tools under human control (Leonardi 2012; Orlikowski 2010; Pickering 1993).

What is the reason behind the discrepancy between practice and academia? The CarCorp executive hinted that they had a reputation to protect as well as having laws to obey. Moreover, they had observed that programmed aggressive behaviour has been met with skepticism by the general public prompting them to take a passive approach. Respondents from public sector were likewise skeptical of programming aggressive behaviour. We could see the main reason being as they strive to improve public safety. Autonomous cars offer a way to minimize the risks associated with human driving, prompting a favor towards the passive approach. Again, we see that there are discrepancies in what is expected of autonomous cars and what the true value really is. Academia might have a slight more edge towards factors not only concerned with safety, such as increasing the socio-economic benefits or lowering the ecological footprint of transport. As such, they could arguably be more visionary in seeing that laws and regulations can be altered in the long run to house the technology they promote. Practitioners on the other hand are more bound to laws and regulations that hinders how they are able to use and develop technology.

If we take a general approach that would be true for all respondents, we could identify that autonomous cars should be reactive to its environment. This broad definition could thus be classified as having agency in the AI sense, where an actor senses and act on its environment (Russel & Norvig 2016). However, it is not clear that it has a specific task in the AI sense. We would argue that this is where it gets problematic as far as the AI definition goes as it entails that the autonomous car has a clear goal which it acts to fulfil (ibid). On the same note, how does one prioritize what the most important task is for an autonomous car? Is the autonomous car to be on time, not hit people on the way to its destination, be polite, drive comfortably for the passengers or obeying formal traffic rules? How do you rank what the most important is amongst these? It may not be viable to have them replicate human behaviour in this regard as we would still argue that they will not be able to distinguish between “good” and “bad” behaviour in contingent situations as previously discussed. In addition, we humans tend to alter our goals as we go along to further problematizes if technology is to mimic human behaviour, to single one out as the most important might prove troublesome for that reason.

In accordance to our empirical data, other reports have discussed the limitations regarding the use of AI in autonomous cars and specifically its inability to match human-like judgement in complex situations that involves human intentionality. KPMG and CAR (2012, p. 12) states that:

> *“So far, the fusion of available sensors and artificial intelligence is not capable of “seeing” and understanding the vehicle’s surroundings as accurately as a human being can. Humans use a combination of stored memories and sensory input to interpret events as they occur and anticipate likely scenarios. For example, if a ball were to roll onto a road, a human might expect that a child could follow. Artificial intelligence cannot yet provide that level of inferential thinking, nor can it communicate in real time with the environment.“*

Another issue in respect to the AI definition of agency is that AI is seemingly absent from how autonomous cars operate. That is, it is formally not involved in the decision making as identified in our empirical data. It is quite contradictory for something that is proposedly autonomous that one must know how they handle every situation in advance. This is because the very notion of something autonomous is that it should handle situations in its own right (Winner 1978). This gets further problematic as autonomous cars are to operate in an environment that will arguably provide unique ad hoc situations, providing a possible contradiction between what is expected of them and what is likely to be their application in the real world. While this might be a semantic reflection, it is no less an important one, as the very notion of them being autonomous would, if being true to the concept of AI, incorporate it in the decision making. This further puts the favor for more of an IS definition of agency in relation to autonomous cars, where the technology is under human control. But as stated by Rose et al. (2005) and Andersen et al. (2016), can we have control over technology at all times?

5.9 Limited autonomous agency

A topic for discussion and interest for a substantial amount of time has been whether humans actually have control over their creations (Winner 1978). This can be observed in the works of fiction and popular culture, like Mary Shelley's Frankenstein, 2001: A Space Odyssey, Terminator, The Matrix and I Robot. All of these revolve around how humans lose control over their developed technologies. These thoughts can also be observed in the works of information infrastructures (cf. Ciborra 2001) where it is considered close to pointless in trying to control technology as it seemingly has a will of its own. Winner (1978) addresses this around the mention of mastery, and that we are experiencing a loss of control when our ability to understand the bigger picture blurs as systems and technologies becomes far more complex. The author claims that it thus would be absurd to state that technology have agency on the basis of experiencing a loss of control that is not attributed to technological agency, but to an increase in overall complexity. Doing so would be accepting technological determinism, disregarding that humans are the ones that create and apply technology. These ideas are reflected in the way material agency is treated in the field of IS as mentioned before as well as contemporary philosophy (Floridi 2014) where technology merely is seen as tools used by humans.

What we are seeing that can be rather confusing is when technology proposedly is *designed to be autonomous* as in the scope of this study. As such, technology is created to follow human instructions, but with autonomous cars, in order to follow human instructions, it *should not follow human instructions*. That is, in order for the technology to fulfil the goal we designed it to do, we must let go of control and assign some extent of free will to the technology. This idea invokes a huge amount of questions and people often end up in ethical dilemmas, such as the trolley problem (cf. Thomson 1985), where the absurd of having a machine choose between who lives and who dies is presented (cf. MIT 2017). These ideas can be directly related to the concept of mastery and subsequently giving up control to technology. It is perplexing as we find ourselves in a contradiction; supposedly we create a technology that is to obey our command, but in order for it to function, it must to some extent disobey our commands as it has to act on its own devices.

But is this really the case in autonomous cars? This would mean that the technology itself would have agency or at least simulate agency in social interactions. This study shows, however, that this obviously is not the case, at least not in practice. Technology, as identified in our empirical data, is to follow suit of very strict instructions of humans and it is not allowed to use AI to make decisions. The autonomous part of autonomous cars is thus at the moment not very autonomous in the classical sense of the word. Their room to act on their own is limited extensively to geographical locations and certain traffic situations, where only other vehicles are present. In addition, future plans for autonomous cars often involve some sort of enveloping the infrastructure where environmental complexity is reduced in one way or another.

As such, the need to revise the notion of autonomous agency in IS research (Andersen et al. 2016) proved to be less pressing as we identify the IS definition as being heavily favored by our empirical data and practice (cf. Volvo 2015) as well as evidence by passed laws in practice (cf. Kang 2016; Näringsdepartementet 2016). Furthermore, our empirical data shows that if technology is to operate autonomously in close relation to humans, it is doing so under heavily constricted forms, which again would favor the IS definition of agency where technology is subject of human control.

However, we could also distinguish that at times, autonomous cars were described as not being under *direct* human control since they in fact were allowed to be operating autonomously under very strict conditions where the level of environmental complexity is low. In light of this, we propose the term "*limited autonomous agency*" to reflect how agency is attributed to autonomous cars. That is, they are autonomous but limited in the sense that they operate under very controlled conditions in environments that have low levels of complexity. The control is always in an arm's reach of humans, either directly or by third party. As previously discussed, this can be seen in the empirical data and in practice where autonomous cars are only permitted to be autonomous when their manufacturers can guarantee that they are in control over the technology.

5.10 Implications for theory and practice

The study adds to the theoretical body of sociomateriality in IS research regarding autonomous agency. Previously, technological agency have mainly been described as being reliant on humans to function. As technology moves towards full automation, this definition have been found not applicable (Andersen et al. 2016). However, the results of this study show that the prior definition is mainly still viable as technology in practice is still under human control. The main theoretical contribution is that the study identifies that the technology is not under *direct* human control, it is given leeway to be autonomous under very strict conditions. If something unforeseen is to happen that the technology can not address itself, then control is retaken by humans. As such, we propose the term *"limited autonomous agency"* to reflect how technology can be seen as autonomous but still in an arm's reach of human control.

A main concern regarding autonomous cars is that the practical implications of having them interact with humans in a social context have largely been unexplored. The study gives insight into what these problems might be as well as providing some suggestions in how to deal with these. Further, it provides a holistic view of how different stakeholders perceive how the introduction of autonomous cars in a social context best is addressed, providing valuable insight in understanding and eventually bridging these differences.

5.11 Limitations

The locus of this study is of autonomous cars whereas we strive to add to the research of autonomous technology in general when it is applied in the social environment of traffic. We find our results relevant for other autonomous technologies besides autonomous cars as many of the concerns raised are not limited to autonomous cars alone. For instance, it addresses the issues regarding how much control one could assign to technology that is to operate in close human proximity. Moreover, the study takes the position that human intelligence is different from technological intelligence, this is often described as the position of Weak AI (Duffy 2003; Floridi 2014). In effect, the study addresses the discussion if technology should simulate human behaviour and what effects doing so might have. The results would likely be different in having an alternative view of what is considered as intelligent as our main ideas revolves around the idea that only humans possesses “true” intelligence (Floridi 2014). This debate quickly becomes philosophical as researchers cannot agree on what is considered to be intelligent and if something that simulates intelligence is to be considered intelligent (Russel & Norvig 2016). We leave this discussion open as the future of AI is yet to unfold.

Moreover, regarding AI in autonomous cars, we identified that AI is not formally involved in the car’s decision making. This could differ between manufacturers, how much they “dare” to involve AI technology for crucial decisions that involves predicting pedestrian behaviour for instance. Nevertheless, we would still argue that using AI would contradict the laws stating how manufacturers must have control over their technologies at all times as previously discussed. Going forwards, this could be subject of change with further progress in the field of AI. In this case, for AI to be implemented in the decision making of autonomous cars, it must significantly surpass the abilities of humans in terms of safety. It must also interact flawlessly with real human beings in contingent situations as traffic most likely will involve social interactions even in the future since roads are part of the open environment of traffic and society. The technology would have to function to the degree that we can depend on that it does not do anything out of what it is supposed to. And in this case; would they still be subject of human agency or would they then be considered truly autonomous?

Additionally, regarding the method used, the traffic scenarios are mostly concerning autonomous cars in city traffic. Most car manufacturers have yet to encompass these situations. As such, this could have impacted the results since they have yet to develop technologies to manage the situations that occur in the scenarios presented. However, as they strive to broaden the scope of autonomous cars to deal with situations analog to the ones described by our scenarios, we find it useful to explore how to deal with these situations as they will be of interest in the near future.

Further, a common critique towards using the snowballing technique is that researchers may risk getting respondents that all share the same ideas as they are part of the same network (Biernacki & Waldorf 1981). We would argue that in our case, our results show that we avoided this pitfall since all respondents had quite different views of the subject at hand. This might be because their network is very broad and seemingly involves people with contrasting opinions and interests.

One may also ask why we chose not to involve regular civilians in the study as they too will be affected by the introduction of autonomous cars. We chose not to include civilians as they are not formally involved in the development of autonomous cars and are arguably not that familiar with the technology or its applications. Such a study would focus on matters of public perception of autonomous cars rather than the development of autonomous cars which puts it outside the scope of this study. Moreover, such studies have been made previously (cf. Kyriakidis et al. 2015) and we thus found including civilians not relevant for the study.

5.12 Suggestions for further research

As the field of autonomous cars is widely unexplored, there is a plethora of subjects for further research. We will propose some subjects that we find to be relevant in relation to the scope of this study:

Given that autonomous cars will most likely house our roads in a near future, situations will occur when liability and responsibility problems eventually have to be considered. That is, if autonomous cars are to operate without human fallback, whom is actually going to be held responsible when an autonomous car causes an accident? Is it the firm that produced the car, one of the engineers that programmed the car's functions, a high executive or someone else?

Further, we identified that stating the overall goal of a vehicle was proved troublesome. As most autonomous cars operate as hierarchical state machines, they operate under the conditions that some goals are superordinate others, but how does one choose which is the most prevalent?

Finally, we excluded cyclists and only referred to pedestrians and human drivers when referring to human actors in traffic. During the empirical data collection, however, cyclists were mentioned as being hard to deal with and the interaction between cyclists and autonomous cars are thus just as important to study, especially when autonomous cars are to operate in a social environment that houses both cyclists and pedestrians as well as vehicles. Are there actual differences between cyclists and vehicles in terms of how their movement can be predicted? Could cyclists be easier to predict than pedestrians as they are bound to the physical laws of how bikes can turn and stop?

6. Conclusions

The aim of this study was to explore how agency is attributed to autonomous cars as they are to be introduced in the social environment of traffic. We identified that the prevalent method of introduction involved adapting the environment to them. This proved, however, to be problematic by both the empirical data and theory as infrastructures are not easily altered. Moreover, some respondents from academia attributed human-like agency to technology where autonomous cars were given mandate to cooperate with humans and thus simulate social agency. On the contrary, practitioners, politicians and respondents from public sector defined autonomous cars as having agency in classical IS-fashion, where technology is seen as tools under human control. Here, technology acts passively and merely reacts to its environment, not cooperating with humans to the same extent and in effect not enacting social agency. The latter definition proved to be most prevalent in the study as well as the real world where technology in use as well as laws and regulations reflect this notion. However, the autonomous cars were under very strict conditions operating fully autonomously which made the prior definition not comprehensible enough. As such, we propose the term "*limited autonomous agency*" that reflects how agency is attributed to autonomous cars as they are operating autonomously while always being in arm's reach of human control.

References

Accidentsketch (2017). *Create a perfect sketch of your accident*. http://draw.accidentsketch.com/ [2017-03-15].

Andersen, J.V., Lindberg, A., Lindgren, R. & Selander, L. (2016). Algorithmic Agency in Information Systems: Research Opportunities for Data Analytics of Digital Traces. In *49th Hawaii International Conference on System Sciences (HICSS)*, Koloa, Hawaii, USA, January 5-8, 2016, 4597-4605.

Anderson, J.M., Kalra, N., Stanley, K.D., Sorensen, P., Samaras, C. & Oluwatola, O.A. (2016). *Autonomous Vehicle Technology: A Guide for Policymakers*. 2nd edn. Santa Monica, CA: Rand Corporation.

Andersson, B.E. (2001). *Som Man Frågar får Man Svar - En Introduktion i Interju- och Enkätteknik*. Stockholm: Prisma ePan.

Becker, J., Helmle, M. & Pink, O. (2017). "System Architecture and Safety Requirements for Automated Driving". In D. Watzenig & M. Horn (Eds.). (2017). *Automated Driving: Safer and More Efficient Future Driving*. Cham, Switzerland: Springer International Publishing, 265-283.

Beiker, S. (2014). "History and Status of Automated Driving in the United States". In G. Meyer & S. Beiker (Eds.). (2014). *Road Vehicle Automation*. Cham, Switzerland: Springer International Publishing, 61-70.

Bengler, K., Dietmayer, K., Färber, B., Maurer, M., Stiller, C. & Winner, H. (2014). Three Decades of Driver Assistance Systems: Review and Future Perspectives. *IEEE Intelligent Transportation Systems Magazine*, 6(4), 6-22.

Biernacki, P. & Waldorf, D. (1981). Snowball Sampling: Problems and Techniques of Chain Referral Sampling. *Sociological Methods & Research*, 10(2), 141-163.

Braun, V. & Clarke, V. (2006). Using Thematic Analysis in Psychology. *Qualitative Research in Psychology,* 3(2), 77–101.

Brewster, S. (2016). *Uber Starts Self-Driving Car Pickups in Pittsburgh*. https://techcrunch.com/2016/09/14/1386711/ [2017-02-15].

Brown, B. & Laurier, E. (2017). The Trouble with Autopilots: Assisted and Autonomous Driving on the Social Road. *Proceedings of CHI 2017*.

Brustoloni, J.C. (1991). *Autonomous Agents: Characterization and Requirements.* Carnegie Mellon Technical Report CMU-CS-91-204. Carnegie Mellon University.

Bryman, A. (2012). *Social Research Methods*. 4th edn. New York: Oxford University Press.

Ciborra, C.U. (2001). "A Critical Review of the Literature on the Management of Corporate Information Infrastructure". In C.U. Ciborra et al. (Ed.). (2001). *From Control to Drift. The Dynamics of Corporate Information Infrastructures*. New York: Oxford University Press, 15-40.

Ciborra, C.U. & Hanseth, O. (2001). "Introduction". In C.U. Ciborra et al. (Ed.). (2001). *From Control to Drift. The Dynamics of Corporate Information Infrastructures*. New York: Oxford University Press, 1-11.

Conner-Simons, A. & Gordon, R. (2016). *Teaching Machines to Predict the Future*. http://news.mit.edu/2016/teaching-machines-to-predict-the-future-0621 [2017-04-16].

Duffy, B.R. (2003). Anthropomorphism and the Social Robot. *Robotics and Autonomous Systems*, 42(3), 177-190.

Floridi, L. (2014). *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*. Oxford: Oxford University Press.

Fraedrich, E., Beiker, S. & Lenz, B. (2015). Transition Pathways to Fully Automated Driving and its Implications for the Sociotechnical System of Automobility. *European Journal of Futures Research*, 3(1), 1-11.

Färber, B. (2016). "Communication and Communication Problems Between Autonomous Vehicles and Human Drivers". In M. Maurer, J.C. Gerdes, B. Lenz & H. Winner (Eds.). (2016). *Autonomous Driving: Technical, Legal and Social Aspects*. Berlin Heidelberg: Springer, 125-144.

Gao, J., Lee, J.D. & Zhang, Y. (2006). A Dynamic Model of Interaction Between Reliance on Automation and Cooperation in Multi-Operator Multi-Automation Situations. *International Journal of Industrial Ergonomics*, 36(5), 511-526.

Gomes, L. (2014). *Hidden Obstacles for Google's Self-Driving Cars.* https://www.technologyreview.com/s/530276/hidden-obstacles-for-googles-self-driving-cars/ [2017-02-15].

Gray, R. (2014). *Google teaches its driverless cars to be more AGGRESSIVE: Vehicles now nudge into traffic to compete with pushy drivers.* http://www.dailymail.co.uk/sciencetech/article-2850530/Google-teaches-driverless-cars-AGGRESSIVE-Vehicles-nudge-traffic-compete-pushy-drivers.html [2017-05-17].

Holmberg, P-E., Collado, M., Sarasini, S. & Williander, M. (2016). *Mobility as a Service - MaaS: Describing the Framework*. Victoria Swedish ICT. Final Report MaaS Framework, 2(54).

Johansson, I. (2017). *Så Ska Göteborg Anpassas för Självkörande Bilar*. http://www.gp.se/nyheter/ekonomi/s%C3%A5-ska-g%C3%B6teborg-anpassas-f%C3%B6r-sj%C3%A4lvk%C3%B6rande-bilar-1.4195828 [2017-03-16].

Juhlin, O. (2010). *Social Media on the Road: The Future of Car Based Computing*. London: Springer Verlag.

Kang, C. (2016). *Self-Driving Cars Gain Powerful Ally: The Government*. https://www.nytimes.com/2016/09/20/technology/self-driving-cars-guidelines.html?_r=0 [2017-03-02].

Kaptelinin, V. & Nardi, B.A. (2006). *Acting with Technology: Activity Theory and Interaction Design.* Cambridge, Massachusetts: MIT Press.

Kaznov, V., Svahn, J., Roos, P., Asplund, F., Behere, S. & Törngren, M. (2017). "Architecture and Safety for Autonomous Heavy Vehicles: ARCHER". In D. Watzenig & M. Horn (Eds.). (2017). *Automated Driving: Safer and More Efficient Future Driving*. Cham, Switzerland: Springer International Publishing, 571-581.

Kennedy, M. (2016). *100 Familjer Får Testa Självkörande Volvo*. http://www.gp.se/nyheter/ekonomi/100-familjer-f%C3%A5r-testa-sj%C3%A4lvk%C3%B6rande-volvo-1.3798391 [2017-02-15].

Klein, H.K. & Myers, M.D. (1999). A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems. *MIS Quarterly*, 23(1), 67-93.

KPMG and Center for Automotive Research (CAR). (2012). *Self-driving cars: The next revolution.* https://home.kpmg.com/be/en/home/insights/2012/08/self-driving-cars-the-next-revolution.html#Link1 [2017-04-17].

Kröger, F. (2016). "Automated Driving in Its Social, Historical and Cultural Contexts". In M. Maurer, J.C. Gerdes, B. Lenz & H. Winner (Eds.). (2016). *Autonomous Driving: Technical, Legal and Social Aspects*. Berlin Heidelberg: Springer, 41-68.

Kurt, A. & Özgüner, Ü. (2013). Hierarchical Finite State Machines for Autonomous Mobile Systems. *Control Engineering Practice,* 21(2), 184-194.

Kyriakidis, M., Happee, R. & de Winter, J.C.F. (2015). Public Opinion on Automated Driving: Results of an international Questionnaire Among 5000 Respondents. *Transportation Research Part F* 32, 127-140.

Lee, J.D. & See, K.A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors*, 46(1), 50-80.

Leitner, A., Ochs, T., Bulwahn, L & Watzenig, D. (2017). "Open Dependable Power Computing Platform for Automated Driving". In D. Watzenig & M. Horn (Eds.). (2017). *Automated Driving: Safer and More Efficient Future Driving*. Cham, Switzerland: Springer International Publishing, 353-367.

Leonardi, P.M. (2012). “Materiality, Sociomateriality, and Socio-Technical Systems: What Do These Terms Mean? How Are They Related? Do We Need Them?” In P.M. Leonardi, B.A. Nardi & J. Kallinikos (Eds.). (2012). *Materiality and Organizing: Social Interaction in a Technological World*. Oxford: Oxford University Press, 25-48.

Lev-Ram, M. (2013). *Teaching IBM’s Watson the meaning of ‘OMG’.* http://fortune.com/2013/01/07/teaching-ibms-watson-the-meaning-of-omg/?iid=sr-link1 [2017-04-19].

Maes, P. (1995). Artificial Life Meets Entertainment: Lifelike Autonomous Agents. *Communications of the ACM,* 38(11), 108-114.

Maurer, M. (2016). “Introduction”. In M. Maurer, J.C. Gerdes, B. Lenz & H. Winner (Eds.). (2016). *Autonomous Driving: Technical, Legal and Social Aspects*. Berlin Heidelberg: Springer, 1-7.

MIT (2017). *Massachusetts Institute of Technology media lab: Moral Machine.* http://moralmachine.mit.edu/ [2017-05-22].

Myers, M.D. (1997). Qualitative Research in Information Systems. *MIS Quarterly,* 21(2), 241-242.

Myers, M. D. & Newman, M. (2007). The Qualitative Interview in IS Research: Examining the Craft. *Information and Organization*, 17(1), 2-26.

Nambisan, S. & Sawhney, M. (2007). A Buyer's Guide to the Innovation Bazaar. *Harvard Business Review*, June 2007, 109-118.

NHTSA. (2013). *National Highway Traffic Safety Administration: Preliminary Statement of Policy Concerning Automated Vehicles*. United States Department of Transportation.

NHTSA. (2016). *National Highway Traffic Safety Administration: Federal Automated Vehicles Policy Accelerating the Next Revolution in Roadway Safety*. United States Department of Transportation.

Näringsdepartementet. (2016). *Vägen till Självkörande Fordon - Försöksverksamhet*. (SOU 2016:28). Stockholm: Regeringskansliet.

Orlikowski, W.J. & Iacono, C.S. (2001). Research Commentary: Desperately Seeking the "IT" in IT Research - A Call to Theorizing the IT Artifact. *Information Systems Research*, 12(2), 121-134.

Orlikowski, W.J. (2007). Sociomaterial Practices: Exploring Technology at Work. *Organization studies*, 28(9), 1435-1448.

Orlikowski, W.J. (2010). The Sociomateriality of Organisational Life: Considering Technology in Management Research. *Cambridge Journal of Economics*, 34(1), 125-141.

Patel, R. & Davidson, B. (2011). *Forskningsmetodikens Grunder: Att Planera, Genomföra och Rapportera en Undersökning*. Lund: Studentlitteratur.

Pickering, A. (1993). The Mangle of Practice: Agency and Emergence in the Sociology of Science. *American Journal of Sociology*, 99(3), 559-589.

Pickering, A. (2001). "Practice and Posthumanism: Social Theory and a History of Agency". In. T.R. Schatzki, K.K. Cetina & E. von Savigny (Eds.). (2001). *The Practice Turn in Contemporary Theory*. London: Routledge, 172-183.

Rose, J., Jones, M. & Truex, D. (2005). Socio-Theoretic Accounts of IS: The Problem of Agency. *Scandinavian Journal of Information Systems*, 17(1), 133-152.

Russel, S. & Norvig, P. (2016). *Artificial Intelligence: A Modern Approach*. 3rd Global edn. Upper Saddle River, New Jersey: Pearson Education Limited.

SAE J3016. (2016). *Surface Vehicle Recommended Practise: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. Standard: J3016_201609.

Schoettle, B. & Sivak, M. (2015). *Potential Impact of Self-driving Vehicles on Household Vehicle Demand and Usage* (Report No. UMTRI-2015-3). Ann Arbor: University of Michigan Transportation Research Institute.

Seppelt, B.D. & Lee, J.D. (2007). Making Adaptive Cruise Control (ACC) Limits Visible. *International Journal of Human - Computer Studies*, 65(3), 192-205.

Sheridan, T.B. (2016). Human–Robot Interaction: Status and Challenges. *Human Factors: The Journal of Human Factors and Ergonomics Society*, 58(4), 525-532.

Stenovec, T. (2015). *Google has gotten incredibly good at predicting traffic — here's how.* http://www.businessinsider.com/how-google-maps-knows-about-traffic-2015-11?r=US&IR=T&IR=T [2017-04-16].

Tesla. (2017). *Autopilot | Tesla*. https://www.tesla.com/autopilot?redirect=no [2017-02-16].

Thomson, J.J. (1985). The Trolley Problem. *The Yale Law Journal*, 94(6), 1395-1415.

Villasenor, J. (2014). Products Liability and Driverless Cars: Issues and Guiding Principles for legislation. *Brookings Institution*, 2-25.

Volvo. (2015). *US urged to establish nationwide Federal guidelines for autonomous driving*. https://www.media.volvocars.com/global/en-gb/media/pressreleases/167975/us-urged-to-establish-nationwide-federal-guidelines-for-autonomous-driving [2017-05-22].

Volvo. (2017). *Autopilot | Self Driving Volvos by 2017*. http://www.volvocars.com/au/about/innovations/intellisafe/autopilot [2017-02-13].

Walsham, G. (2006). Doing Interpretive Research. *European Journal of Information Systems*, 15(3), 320-330.

Walsham, G. (2014). Empiricism in Interpretive IS Research: A Response to Stahl. *European Journal of Information Systems*, 23(1), 12-16.

Webster, J. & Watson, R.T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, 26(2), 13-23.

Widerberg, K. (2002). *Kvalitativ Forskning i Praktiken*. Lund: Studentlitteratur.

Winner, L. (1978). *Autonomous Technology: Technics-out-of-Control as a Theme in Political Thought.* Paperback edition. Cambridge, Mass: The MIT Press.

Winner, H. & Wachenfeld, W. (2016). "Effects of Autonomous Driving on the Vehicle Concept". In M. Maurer, J.C. Gerdes, B. Lenz & H. Winner (Eds.). (2016). *Autonomous Driving: Technical, Legal and Social Aspects*. Berlin Heidelberg: Springer, 255-275.

Appendix 1: Themes in coding

Theme	**Main occurrence**
Transition period	General
Hard to know what the real value is	General
V2V communication	Scenario 1, 4
Resemble humans	Scenario 1, 2, 3
Understanding other vehicles	Scenario 1, 4
Understanding pedestrians	Scenario 2, 4
Understanding humans	Scenario 3, 4
Indicating if cars are autonomous	Scenario 2, 3
Over-reliance / over confidence to autonomous cars	Scenario 2
Probing behaviour	Scenario 1, 2, 4
Third party control	Scenario 1, 2, 3, 4
Fallback-plan	Scenario 1, 2, 3, 4
Infrastructural requirements	Scenario 2, 3, 4
Altering the infrastructure	Scenario 1, 2, 3, 4

Appendix 2: Merged themes

Theme	Merged into
Transition period, Hard to know what the real value is, V2V communication	General observations
Resemble humans, Probing behaviour	Designing passive or aggressive behaviour
Making others know they are autonomous, Over-reliance / over-confidence to autonomous cars	Indicating if cars are autonomous
Understanding other vehicles, Understanding pedestrians, Understanding humans	Understanding humans, pedestrians and other vehicles
Third party control, Fallback-plan	Third party control
Infrastructural requirements, Altering the infrastructure	Altering the infrastructure

Appendix 3: Interview questions

General questions

- How will the transition towards autonomous cars look like?
- Is traffic a social environment?
- What is the actual value of autonomous cars?

Scenario 1: Autonomous car and human driver interaction

- What is the best way to interact between the autonomous car and the human driver?
- What different options does it have?
- How can the autonomous car communicate with other “regular” cars at a distance?
- How can the autonomous car and the human driver cooperate?

Scenario 2: Autonomous car and pedestrian interaction

- Should the autonomous car show that it is autonomous?
- Should it use its “body language” in some way?
- How does it communicate with pedestrians?
- How can the autonomous car cooperate with pedestrians?

Scenario 3: Autonomous car and police officer interaction, unique situations

- How can the autonomous car understand human intentions?
- How much leeway/control can one give to the autonomous car in unique complex situations?
- What does the autonomous car do if it approaches a situation it does not understand?

Scenario 4: Autonomous car interacting with both human drivers and pedestrians

- How does the autonomous car interact with both pedestrians and human drivers at the same time?
- How can autonomous cars manage the concept of traffic flow?