THESIS FOR THE DEGREE OF LICENTIATE OF ENGINEERING

The automation effect

Investigating factors that influence the driver response process in a safety-relevant event during assisted driving and after unsupervised automation

LINDA PIPKORN

Department of Mechanics and Maritime Sciences

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2020

The automation effect
Investigating factors that influence the driver response process in a safety-relevant event during assisted driving and after unsupervised automation
LINDA PIPKORN

Technical report no 2020:10

Department of Mechanics and Maritime Sciences
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Cover:
A test vehicle follows a lead vehicle on a rural road test track. Suddenly, a conflict object (a garbage bag) appears in the lane. The lead vehicle cuts out of lane to avoid the object, causing the object to become fully visible to the driver in the test vehicle. The driver in the test vehicle needs to act to avoid a crash.
Picture created with *Gravit Designer* (https://designer.gravit.io/).

Chalmers Reproservice
Gothenburg, Sweden 2020

Abstract

Introduction: Safe vehicle automation can be achieved through a detailed understanding of drivers' ability to respond to a safety-relevant event after a period of automated driving. For instance, there is a need to understand in which scenarios automation effects are present (e.g. delayed response, degraded driving performance, crashing). Further, there is a need to identify specific factors (e.g. test environment, system-prompts, hands-on-wheel requirement, automation duration) that contribute to or prevent these automation effects. **Objectives:** The aim of this thesis is to investigate factors that influence: (a) automation effects in a non-prompted (i.e. absence of warning/notification) safety-relevant event during assisted driving and (b) automation aftereffects (i.e. automation effects specifically occurring after automation has been deactivated) in a prompted safety-relevant event during unsupervised automation. **Method:** Two Wizard-of-Oz test-track experiments were performed in order to investigate the driver response process in safety-relevant events. In experiment 1, the drivers were required to supervise (with or without a hands-on-wheel requirement) an assisted driving system, and then respond to a safety-relevant event that was not prompted by the system. In experiment 2, the drivers drove manually (baseline) and with an unsupervised automation system (a short and a long duration) before encountering a safety-relevant event. The automation system prompted (issued a take-over request) the driver to resume manual driving shortly before the safety-relevant event became visible. **Results:** In experiment 1, one third of the drivers responded late, or did not act at all, and crashed in the non-prompted safety-relevant event. In fact, the drivers crashed to the same extent and responded similarly independent of if they supervised the assisted driving system with or without hands on the wheel. In experiment 2, all drivers resumed manual control and did not collide in the safety-relevant event, both after a short and a long automation duration. All drivers showed a similar response and driving performance in the safety-relevant event for both long and short automation duration as well as in the manual baseline. **Discussion:** A hands-on-wheel requirement was not found to prevent late response or crashing in a non-prompted safety-relevant event encountered during assisted driving. More work is needed to understand the potential safety-benefits of a hands-on-wheel requirement in other types of conflicts and for driver distractions. The finding of minor automation aftereffects in experiment 2 contrasts to previous driving simulator studies. The reason may be the different test environments but is more likely due to different timings for prompting the drivers to resume manual control in relation to when the safety-relevant event became visible. **Conclusions:** Safe vehicle automation, including both assisted and unsupervised automation, can be achieved in a realistic environment (test track) for most drivers. However, assisted driving in combination with a non-prompted safety-relevant event, can be detrimental for safety, since some drivers may not understand the need to respond to avoid a crash. In fact, a hands-on-wheel requirement did not result in earlier steering responses nor did it prevent drivers from crashing. Thus, more work is needed to understand how to make sure drivers understand the need to respond in non-prompted safety-relevant events during assisted driving. In fact, it seems that when automation has matured to a level when it can prompt the drivers (i.e. unsupervised automation that can issue a take-over request) prior to a safety-relevant event becomes visible, drivers are able to safely resume manual control and perform similar as after an extended period of manual driving. Such safe driving performance seems to be independent of automation durations below 15 minutes.

Key words: automated driving, driver behaviour, driver response, driving performance, take over, test track, response process, automation

Acknowledgements

This research was mainly supported by the European Project L3Pilot, grant number 723051, and partly by the Swedish FFI project ADEST, grant number 2014-06012, within the Drive Me programme.

Firstly, I owe my three supervisors Marco, Trent, and Emma many thanks. I thank Marco for his support no matter what. I thank Trent for his experience and enthusiasm, and Emma for always being there. All of you have helped me grow in different ways.

Secondly, I would like to thank my colleagues at Chalmers. A special thanks to Thomas for being supportive and caring, and to Giulio for introducing me to the topic of human factors in driving a few years ago.

Thirdly, I would like to thank everyone at Volvo Cars who has in any way been involved in my work. Thanks for data collection on ASTA, thanks for interesting lunch discussions and thanks for making me feel welcome. You know who you are.

Finally, what would I have done without my sports, my friends and family? I owe many thanks to everyone on the Tennis court and in the Padel box for giving me positive energy and something else to think of! Thank you to my friends and family for encouragement and always being there to listen when needed.

Peter, thank you for all the love and endless support. You're the best!

Contents

1 Introduction

At the time of writing this thesis, 1.35 million human lives are lost every year in road traffic (World Health Organization [WHO], 2018). In addition, WHO (2018) reports that road-traffic crashes are the leading cause of death for young people between 5 and 29 years. *Vehicle automation* — technologies introduced to automate parts of the driving task in a passenger car — seems promising to have a positive impact on road safety by reducing the number of road-traffic deaths. The reason is that automated vehicles are expected to perform better than the human driver, potentially avoiding 94% of the crashes that are related to driver-related critical reasons such as recognition errors, decision errors, and performance errors (National Highway Traffic Safety Administration, [NHTSA], 2015).

However, in order to obtain this safety-benefit, vehicle automation needs to be designed to be *safe,* by accommodating the human factors challenges that comes with introducing automation to a human-machine system (Bainbridge, 1983; Lee, Wickens, Liu & Boyle, 2017; Seppelt & Victor, 2016). In practice, introducing automation may result in safer driving in some sense (e.g. by reducing driver workload; De Winter, Happee, Martens & Stanton, 2014), but automation may also induce possibilities for new driver behaviours that may be safety-critical in different ways (e.g. drivers that devote less attention to the forward road; Morando, 2019). In fact, on-market vehicles equipped with automated technologies have already been involved in crashes with deadly outcome (National Transportation Safety Board [NTSB], 2017, 2018, 2019). A common factor in these reports is the lack of driver engagement (e.g. long durations hands-off-wheel times) prior to crashing and that the safety-relevant events were not avoided by the automated systems (i.e. the driver needed to act to avoid crashing because the vehicle could not handle the safety-relevant event).

For the scope of this thesis, *safety* will be measured through investigations of the drivers' ability to start to drive manually and respond to a safety-relevant event after a period of automated driving. This ability will be assessed through investigations of the *driver response process* in the safety-relevant event. The *response process* consists of a *response preparation* phase (i.e. including actions that needs to be performed before the driver can start to drive manually e.g. put hands on the steering wheel) and a *manual intervention and stabilization* phase (i.e. including the manual intervention and driving performance that takes place after the drivers have started driving manually). *Automation effects* are present when parts of the driver response process are unsafe (e.g. delayed response, a degraded driving performance, or crash involvement). Therefore, to achieve safe vehicle automation, these effects need to be mitigated or prevented.

1.1 Human collaboration with automation: potential and challenges

Today, automation is increasingly present in technology. Recent advances in software and hardware enable machines to perform tasks that were traditionally performed by a human operator. Introducing automation to a human-machine system comes with both potential and challenges, which have received attention from human factors researchers within the past four decades (Bainbridge, 1983; Lee, et al., 2017; Onnasch, Wickens, Li, & Manzey, 2014; Parasuraman & Riley, 1997; Seppelt & Victor, 2016). On one hand, automation has the potential to increase human safety and comfort (Lee et al., 2017). Machines can efficiently and with high accuracy handle routine tasks that may

either be difficult or even dangerous for humans to perform. On the other hand, automation can also decrease human safety and comfort. The reason is that automation may fail at times, either due to a software or hardware failure or because the automated system is used in a way that was not intended (i.e. beyond system limitations). When automation fails the consequences can be catastrophic.

A catastrophic outcome may stem from the limitations of a human operator to step in as a problem solver when automation fails (often unexpectedly and silently) after having performed well for a long period (Bainbridge, 1983; Wickens, Hooey, Gore, Sebok, & Koenicke, 2009). The reason is that automation alters the tasks performed by the human operator. That is, instead of being actively engaged in operational control the human operator is often left to passively monitor the automated system which performs the tasks previously performed by the human. Consequently, operators may enter an *out-of-the-loop* state, which in turn is linked to a reduced *situation awareness* (i.e. a reduced awareness of the system and processes) and a loss of manual skills (Endsley, 1995, 2015; Endsley & Kiris, 1995). In addition, the higher the degree of automation the less fit humans may be to perform manual operation when needed (e.g. due to automation failure), after automation have performed some of the tasks for a longer period (Onnasch et al., 2014). To conclude, knowing that humans may have difficulties performing manual operation after automation has performed the tasks, may help in solving the new challenges that are presented when drivers are required to drive manually after some period of automated driving in increasingly automated vehicles (Seppelt & Victor, 2016).

1.2 Manual driving and event-response during assisted driving and after unsupervised automation

Driving relies on the human ability to sense and gather information about the driving environment, attend to and process the relevant or important information, and respond to the numerous possible driving situations, both during routine driving and in safety-relevant events (Bolstad, Cuevas, Wang-Costello, Endsley, & Angell, 2010; Macadam, 2003). The driving task consists of three hierarchical levels of skills and control (Michon, 1985). At the top level (the *strategic* level) the general planning of the trip is executed (e.g. decisions on where to go and how to go there, if an automated function should be used or not). At the middle level (the *tactical* level) decisions are made regarding the manoeuvring control related to the present circumstances (e.g. speed selection, avoiding obstacles, turning, overtaking). Finally, at the lowest level (the *operational level*) the continuous control of the vehicle is performed (e.g. steering, braking, accelerating).

1.2.1 Assisted driving and unsupervised automation

When automation is introduced to a vehicle, the human activities on the operational, tactical and strategic levels may change. Automation is not necessarily "all or none" but exists to different degrees, often called levels of automation, and these degrees may change the human activities in different ways (Parasuraman, Sheridan & Wickens, 2000). There are several definitions of the degree or levels of automation, ranging from complete manual control up to full automation when no human input is required (SAE, 2018; Seppelt & Victor, 2016; Thatcham Research, 2018).

For the scope of this thesis, the focus will be on two types of automation, namely *assisted driving* and *unsupervised automation.* The reason behind the choice of these two definitions is the clear distinction of driver roles and responsibilities, i.e. that the driving is either *shared* when the driver is fully responsible for the driving or *delegated*

when the vehicle is fully responsible for the driving (Seppelt & Victor, 2016). Two motivations behind the need for such a clear distinction is: (a) the observed public confusion about the actual capabilities of current on-market systems (Thatcham Research, 2019) and (b) concerns about the human ability to be prepared to act/intervene when automation fails (Seppelt & Victor, 2016). Another motivation for focusing on assisted driving and unsupervised automation are that these definitions are being adopted by safety rating organisations and insurance institutes (Euro NCAP, 2018; Thatcham Research, 2019) and will therefore likely guide the design of future automated systems.

The most mature automated driving systems on the market today (year 2020) are *assisted driving* systems, while the next type *unsupervised automation* systems are currently being developed. For assisted driving, the vehicle can support the driver with the longitudinal control (accelerating and braking) and the lateral control (steering) by combining Adaptive Cruise Control (ACC) with Lane Centering (LC). However, the driver is always responsible for the driving and is expected to keep the hands on the steering wheel (United Nations Economic Commission for Europe [UNECE], 2017), eyes on the road, and to be prepared to respond to conflict objects and events that can appear at any time (Seppelt & Victor, 2016; Thatcham Research, 2018). The reason why the driver is the main responsible at all times, is the limitations current systems have which require driver action. For example, so-called "cut-in" (i.e. another vehicle enters the lane between the subject vehicle and a lead vehicle) and "cut-out" (i.e. another vehicle leaves the lane of the subject vehicle to avoid a conflict object on the road) scenarios are challenging for current on-market systems (Euro NCAP, 2018). These types of safety-relevant events occur with high frequency in everyday traffic, but still occur rare enough to surprise drivers, especially when critical. Further, these events generally require the driver to act to avoid the crash by steering and/or braking (Euro NCAP, 2018).

Unsupervised automation, on the other hand, can take full responsibility of the driving task (e.g. accelerating, braking, steering and event detection- and response) for certain periods of time. For example, unsupervised automation would need to safely handle the Euro NCAP cut-in- and cut-out scenarios above without driver intervention. The driver is then allowed to disengage from driving (hands off the steering wheel and eyes off the road) and engage in non-driving related activates (NDRA; e.g. playing games). In fact, a United Nations (UN) vehicle regulation for a future low speed *unsupervised automation* system (referred to as "Automated Lane Keeping System") is currently being developed by the *Working Party on Automated/autonomous and Connected Vehicles* (GRVA; UNECE, 2020). In the current version, when a driver needs to resume manual control these systems are required to issue a so-called *transition demand* to notify the driver beforehand for both *planned events* (events known at system activation) and *unplanned events* (unknown at system activation, but assumed likely to happen during driving, e.g. encountering a road-work zone). In addition, for cases when the driver do not respond to this demand by deactivating automation, the system should start a *minimum risk manoeuvre* ("a procedure aimed at minimising risks in traffic, which is automatically performed by the system after a transition demand") earliest 10 s after the transition demand was issued. That is, the vehicle is required to take responsibility for safe driving when the driver is not fit to do so. However, this regulation also assumes that systems can detect and notify the driver about an upcoming safety-relevant event more than 10 s beforehand.

1.2.2 Terminiology and examples regarding the need for manual driving during assisted driving and after unsupervised automation

The process of resuming manual control from unsupervised automation is referred to as a *transition* in the UN vehicle regulation (UNECE, 2020) as well as in the ISO 21959 for *human performance and state in the context of automated driving* (ISO, 2018). This thesis will foremost use the definitions in the ISO 21959 because: (a) the specific emphasis on terms related to driver performance in the context of automated driving and (b) the schematic models for the *transition processes* for driver-initiated and system-initiated transitions. In fact, these models have inspired Figure 1 below. The purpose of Figure 1 is to introduce necessary terminology and to put this terminology into the context of two examples of safety-relevant events that may occur during assisted driving (Scenario 1) and unsupervised automation (Scenario 2).

ISO 21959 uses the word transition for both *assisted driving* and *unsupervised automation,* even if these processes may be fundamentally different. In unsupervised automation, the automated driving system (ADS) prompts the driver to resume manual control (the need to resume manual control is *ADS-prompted*) through a salient notification when needed. Such a notification (referred to as a *Request to Intervene* in the ISO 21959, but the most common term in the literature is a *take-over request* or a TOR), takes place for situations when the limitations of the system is well known. The TOR is marked in Figure 1 (Scenario 2) as the *prompt (TOR).* As shown in Figure 1, this prompt triggers a *driver state transition* which is defined as the "process of transforming the actual driver state (possibly determined by Non-Driving-Related-Activity) to a target driver state suitable to effectively take-over manual control" (ISO, 2018).

In unsupervised automation, this driver state transition means that the driver goes from having no responsibility for safe driving to full responsibility for safe driving (at *automation deactivated;* Figure 1). In contrast, in assisted driving the driving task is shared between the assisted driving system and the driver, since the driver is fully responsible for the driving task but is supported by the assisted driving system. This is shown in Figure 1 (Scenario 1) as a grey bar which is shared between the driver and the system. Thus, a well-defined transition of control does not exist. The driver can either deactivate the system (e.g. by pressing buttons) and drive in manual mode but can also apply steering wheel torque in order to change the vehicle's path while the driver assistance system remains active. However, in both cases the driver is always fully responsible for safe driving.

The need for manual control input during assisted driving can either be *ADS-non-prompted* or ADS-prompted. The manual control need is *ADS-non-prompted* when there is no notification given by the system, but the driver detects the need to respond to a safety-relevant event when a system limitation occurs (e.g. an unexpected object in lane which the system does not detect, or a steering system torque limitation in a curve). Further, the manual control need is ADS-prompted when the driver receives a disengagement notification, or a forward collision warning. *System limitations* (i.e. a system that does not detect an object, because the system was not designed to detect it, and if known it is declared in the user manual) should be differentiated from a *silent-failure* event where the system is not working as intended (e.g. a system that silently fails in a situation it was designed to handle). The need for starting to drive manually and act in a safety-relevant event during assisted driving when no prompt is present

(ADS-non-prompted) and during unsupervised automation when a prompt (TOR) is present (ADS-prompted) are exemplified with Scenario 1 and 2 in Figure 1, respectively.

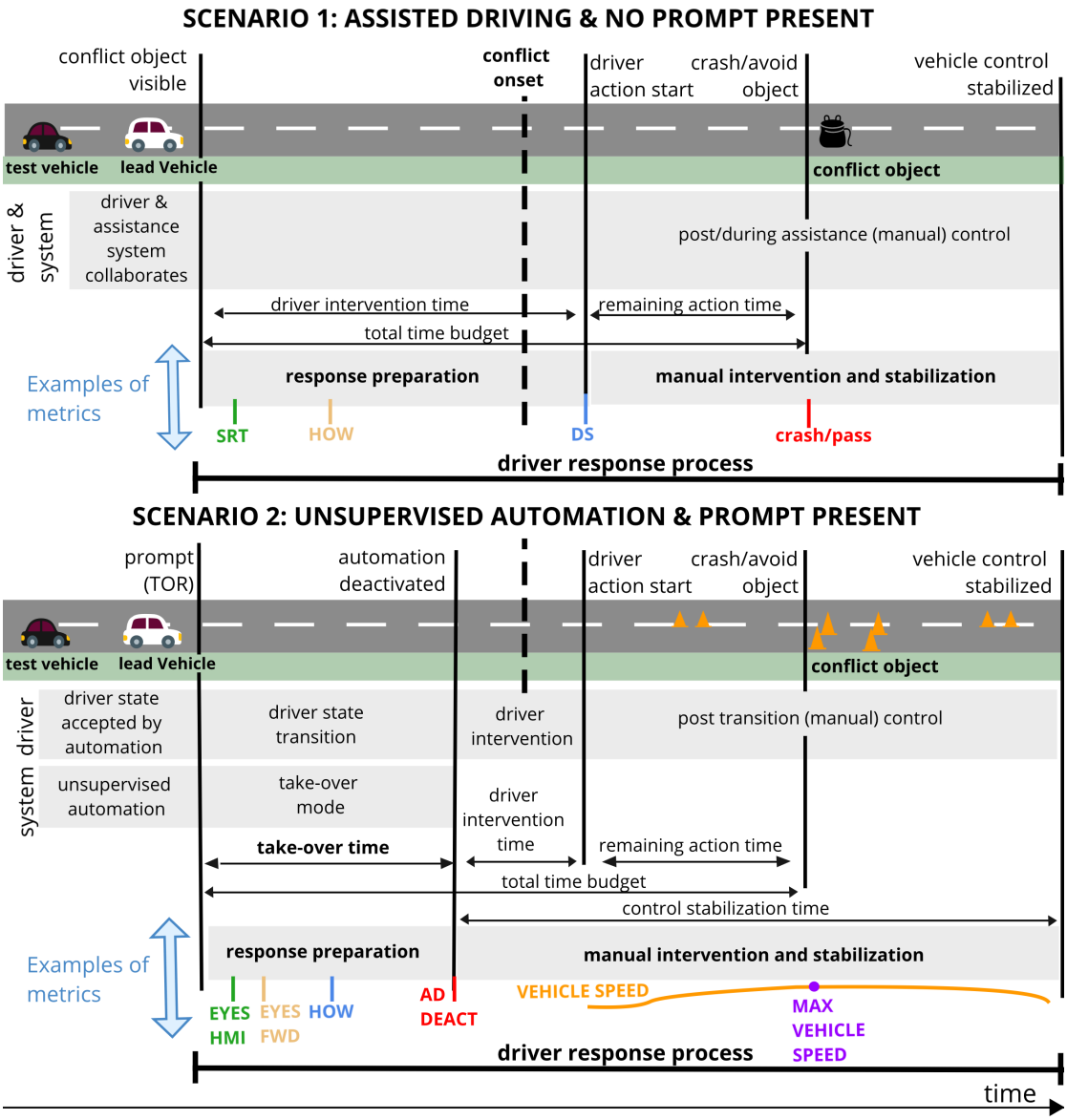

Figure 1 - A representation of the process of driving manually after a period assisted driving (Scenario 1) and unsupervised automation (Scenario 2). The figure defines typical durations of interest for assessing the driver response process (e.g. the take-over time). In addition, the components building up the complete driver response process (the response preparation phase and the manual intervention and stabilization phase) are marked.

Scenario 1

A test vehicle is following a lead vehicle on a rural road. The test vehicle has the assisted driving system engaged which means that the driver is responsible for the driving task but is supported by the assisted driving system (*driver & assistance system collaborates)*. Suddenly, a *conflict object* (garbage bag) becomes visible for a short moment (*conflict object visible)* and can be detected by an attentive driver. A little later, the lead vehicle performs a cut-out and fully reveals the conflict object to the driver (*conflict onset*) who needs to perform an evasive steering manoeuvre (*driver action start* marks the start of this driver intervention) to avoid crashing with the object since the assisted driving system did not detect the object and was not performing any evasive steering manoeuvre. Thus, for a successful driver response process the driver needs to:

(a) understand the need to act without being prompted by the system and, (b) complete the actions in the response preparation phase (e.g. put hands on wheel) to be able to start apply steering torque or deactivate the driver assistance system and, (c) act to avoid crashing. The phase from driver action start is referred to as the *manual intervention and stabilization* phase and includes both the driving performance needed to avoid crashing in case of a conflict object as well as a period of returning to *normal* (i.e. the driver either returns to collaborating with the system or deactivates the system) in case of normal non-critical driving. Note that the cut-out scenario is just one example of a system limitation requiring driver control and many others exist (e.g. see Euro NCAP, 2018).

Scenario 2
A test vehicle is following a lead vehicle on a rural road. This time the test vehicle has the unsupervised automation system activated and the driver is engaged in playing a game on his/her tablet with hands off the wheel and eyes off the road. The unsupervised automation system detects an upcoming road-work zone while the lead vehicle is still in front and prompts the driver to deactivate automation and resume manual control by issuing a TOR. The driver needs to perform a *driver state transition,* i.e. perform some actions in the *response preparation* phase, for example stop playing the game, redirect his/her eyes from the tablet to the human-machine interface (HMI) and/or to the forward road, put the hands on the steering wheel and press two buttons in order to deactivate the unsupervised automation system (*automation deactivated* marks when this is achieved). After having deactivated automation, the lead vehicle changes lane (similar to a cut-out manoeuvre; the *conflict onset*) to avoid colliding with the first two upcoming cones which are part of the road-work zone (the *conflict object*). The driver who is now in manual driving mode needs to start steering (*driver action start*) and then complete the *manual intervention and stabilization* phase by carefully manoeuvring in the road-work zone not to collide with any of the cones and return to a stable manual driving performance (*vehicle control stabilized* marks when this is achieved). For some systems the driver action start may be used to deactivate automation and in such a case the *driver intervention time* would be zero. Note that any number of critical scenarios (anything that could be encountered in manual driving) may at any time occur shortly after the driver has resumed manual control.

1.3 State-of-the-art: automation effects, driver response process, and influencing factors

A growing amount of litereature is concerned with understanding the scenarios (i.e. a combination of e.g. environment, automation type, event type) in which drivers may have a degraded ability to drive manually and respond to a safety-relevant event after a period of automated driving (i.e. when parts of the driver response process are unsafe), as well as the factors that influence this ability (Mole et al., 2019; Eriksson & Stanton, 2017; McDonald et al., 2019; Zhang, De Winter, Varotto, Happee, & Martens, 2019). In other words, there is an interest in understanding in which scenarios automation effects are present, the type and size of the effects, as well as the factors that contribute to or can mitigate such effects. This understanding is important because it can be used to achieve safe vehicle automation by informing the development of, for example: (a) vehicle regulations, (b) countermeasures or system design principles, and (c) driver models to be used in computational simulations to estimate the benefit of those countermeasures (Bärgman, Boda, & Dozza, 2017; McDonald et al., 2019) and for driver monitoring systems (i.e. in-vehicle systems that can infer driver states and actions).

To begin with, understanding the scenarios that are associated with automation effects (i.e. a problem) is important because it gives information about which scenarios are problematic and should be focused on (e.g. a lead-vehicle cut-out scenario encountered in assisted driving). Secondly, understanding the type and size of the effects (e.g. crashing) is important because it informs how severe the problem is. Thirdly, when the scenario and the type of effects are identified, understanding the factors or mechanisms that contribute to these effects (e.g. humans expect automation to act) is needed in order to solve the problem. Finally, understanding the factors that can prevent automation effects (e.g. a hands-on-wheel requirement, restrictions on automation duration) is important because it may help in solving the problem by facilitating an appropriate and safe driver response process and thus prevent crashes.

This Section will summarize what is currently known (the state-of-the-art) about driver response in automation in general, and automation effects (when this response is unsafe) in specific. In addition, this Section will present typical ways of assessing the driver response process in safety-relevant events in automation (to understand presence/absence of automation effects) as well as what is currently known about the influence of specific factors on the response process.

1.3.1 Metrics used to assess the driver response process

In order to understand the scenarios in which automation effects are present as well as the type and size of these effects, means of measuring the driver response process are needed. The driver response to safety-relevant events has traditionally been examined using reaction times (e.g. driver brake or steering reaction times; Young & Stanton, 2007). A reaction time is typically defined as the duration between the onset of a safety-relevant event (e.g. a lead vehicle that starts to brake) and the driver action start (e.g. driver presses the brake pedal). For example, in Scenario 1 shown in Figure 1, the onset of the safety-relevant event could either be defined as the timing of *conflict object visible* for the first time or at the *conflict onset* when the lead vehicle performs the cut-out.

Up until now, the most frequently used metric to assess the driver response process in a safety-relevant event after automation is the so-called *take-over time* or in short *TOT* (Mole et al., 2019; Eriksson & Stanton, 2017; McDonald et al., 2019; Zhang et al., 2019). The TOT is typically defined as the time between when a prompt (TOR) is issued and the time when the driver has deactivated automation by either a button press, or when a certain braking or steering threshold is met (automation deactivated or driver action start; see Figure 1, Scenario 2). As shown in Figure 1, the driver response process consists of more than only a TOT. Recently, it has been proposed that investigations of TOT and the response preparation phase needs to be combined with analyses of the manual driving performance that follows a period of assisted driving or unsupervised automation (Mole et al., 2019; Zeeb, Buchner, & Schrauf, 2016). This is because the factors influencing TOTs may not be the same that influence the driving performance that follows (Zeeb et al., 2016) and some of these factors may only be understood by considering the drivers' steering behaviour after automation (e.g. a less calibrated perceptual-motor control; Mole et al., 2019).

Thus, the analyses of the response preparation phase should be enriched with investigations of the driving performance in the manual intervention and performance phase (Figure 1). However, up until now there is no well-defined way to assess the

success of this manual intervention and performance phase (McDonald et al., 2019; Mole et al., 2019), although it is identified as a control stabilization period in ISO (ISO, 2018). Examples of metrics that have been used to assess the driving performane in the manual intervention and stabilization phase are: conflict outcome, minimum time-to-collision (min TTC), as well as descriptive statistics (mean, maximum, minimum) of longitudinal and lateral accelerations (McDonald et al., 2019).

1.3.2 Automation effects and aftereffects: the influence of automation on the driver response process

It appears that some of the human factors concerns related to automation observed in other domains are also present in certain vehicle automation scenarios. Even lower degrees of automation (e.g. ACC) have been shown to increase driver response times, such as brake reaction times, to safety-relevant events when no prompt to the driver is present and when the response is compared to manual driving (Larsson, Kircher, & Hultgren, 2014; Bianchi Piccinini et al., 2019; Young & Stanton, 2007). Some evidence also indicates that the higher the degree of automation (ACC + Automated steering compared to ACC alone) the poorer the driver response to a safety-relevant event (Strand, Nilsson, Karlsson, & Nilsson, 2014), but evidence of no effect or minor effects of higher degree of assistance on the driver response to safety-relevant events also exist (Larsson et al., 2014; Young & Stanton, 2007). A recent test-track study by Victor et al. (2018) confirmed that drivers may have difficulties responding to a safety-relevant event (i.e. a longitudinal lead vehicle cut-out situation with a stationary conflict object not detected by automation, as previously exemplified in Scenario 1) during assisted driving. In fact, 28% of the drivers, all reporting high trust in automation to act in the conflict scenario, crashed with the stationary conflict object, which was either a garbage bag or a stationary balloon car. The drivers that crashed explained that they did not act, or they acted too late to avoid a crash, because they expected the assisted driving system to avoid the conflict object (Victor et al., 2018).

Evidence of *automation aftereffects* (i.e. automation effects specifically occurring after automation has been deactivated) have also been shown just after unsupervised automation when a TOR is present (Gold, Damböck, Lorenz, & Bengler, 2013; Louw, Merat, & Jamson, 2015; Happee, Gold, Radlmayr, Hergeth, & Bengler, 2017). In addition, some evidence indicates that a longer exposure to automation (i.e. a longer *automation duration*), produces more severe automation aftereffects after a TOR (Bourrelly et al., 2019; Jarosch & Bengler, 2018), whereas at least one other study did not observe any significant automation aftereffects (Feldhütter, Gold, Schneider, & Bengler, 2017).

Notably, few studies have directly investigated how the driver response process may differ between assisted driving and unsupervised automation (McDonald et al., 2019). Thus, little is currently known whether a period of assisted driving or unsupervised automation are associated with similar responses, especially in the same safety-relevant event. The literature gives us reason to be concerned by suggesting that the more we automate, the poorer the human ability to perform manual control after automation seem to become (Onnasch et al., 2014). However, since assisted driving and unsupervised automation differ in whether the drivers are prompted about the need to drive manually or not, this may not always be a general conclusion for both assisted driving and unsupervised automation. Notably, warnings have been found to elicit earlier responses if given well in advance compared to late warnings or no warnings (Lee, McGehee, Brown, & Reyes, 2002). Thus, comparisons of the driver response process between assisted driving and unsupervised automation may both be influenced

by the automation type and the presence of prompts to notify the driver about the need to drive manually.

Finally, all studies mentioned in this Section were performed in driving simulators, except the study by Victor et al. (2018) that was performed on a test track. Thus, the question remains to what extent the above findings of automation effects and aftereffects can be generalised to more realistic contexts (e.g. similar studies but performed on test track or on public road).

1.3.3 The influence of specific factors on the driver response process

Many studies have investigated the influence of specific factors on the TOT. To begin with, TOTs have been found to range between 0.7 s up to 23.8 s (Eriksson & Stanton, 2017; Zhang et al., 2019). The take-over time budget (*Total time budget* in Figure 1) has been pointed out as one of the main influencing factors on TOT: in general the longer the time given to drivers to resume manual control the longer time they seem to take (McDonald et al., 2019). The take-over time budget is typically defined as the TTC at event onset (Figure 1). Additional factors that have been found to influence the TOT are: if the transition-scenario is practiced beforehand, the presence of secondary tasks (especially hand-held tasks) and the presence of prompts to transition. In fact, it seems that TOTs decrease when a prompt (TOR) is present (Zhang et al., 2019), but the influence of prompts on the driver response to safety-relevant events after automation requires additional work (McDonald et al., 2019).

The review by McDonald et al. (2019) concludes that the driving performance in the manual intervention and stabilization phase is significantly influenced by the take-over time budget, secondary task engagement, the modality of the TOR (e.g. visual, auditory, haptic), the driving environment, if a prompt is present or not, repeated exposure, fatigue, trust in automation, and alcohol impairment. Thus, many of these factors influencing driving performance in the manual intervention and stabilization phase are the same as the ones influencing the TOT. This could be because of the relation that exists between: (a) the response preparation phase and (b) the manual intervention and stabilization phase. As the illustrated Scenario 2 in Figure 1 shows: the longer the *response preparation* phase, the shorter *remaining action time* is available. Consequently, in critical scenarios drivers may be forced to act closer to a conflict object and therefore crash or perform an evasive manoeuvre to avoid crashing. This effect of degraded driving performance as a consequence of the time needed for drivers to prepare to act (e.g. positioning the hands to the steering wheel) can be called the *preparation-action-time consequence*. The extent to which the observed automation aftereffects (Section 1.3.2) may be influenced by the preparation-action-time consequence (a timing issue) rather than a human (cognitive) mechanism (e.g. drivers that are less aware of the surroundings because of automation) is currently unknown.

Whereas factors such as the presence of secondary tasks and TOR modality have received a lot of attention (McDonald et al., 2019), some potentially influential factors have not been focused on to the same extent. One example of such a factor is the presence of a hands-on-wheel requirement (McDonald et al., 2019). For drivers to supervise an assisted driving system is today required by law in Europe (UNECE, 2017), whereas future unsupervised automation system may allow drivers to remove their hands from the steering wheel and under some circumstances disengage from the driving task (UNECE, 2020). At least two studies have explicitly studied the influence of permitting hands-off intervals at either 10 s or 120 s on the driver response process during automated driving when a safety-relevant event was preceded by a TOR

(Naujoks et al., 2015, 2017). Naujoks et al. (2015) found that all drivers responded appropriately and with similar brake response times in a longitudinal scenario when encountering a suddenly appearing stationary vehicle for both permitted hands-off durations (10 s vs 120 s). Further, Naujoks et al. (2017) found that drivers also responded similarly independent on the permitted hands-off intervals in a lateral lane-drift scenario. Notably, both studies mention that most drivers did keep hands on the steering wheel in both conditions even if allowed to have hands off, which may have explained the similar response. Another factor that have received little attention but may influence the driver response process is the conflict object type used in the safety-relevant events (McDonald et al., 2019). The conflict object type may influence the driver response process because of different saliency that influences detection and perception (driver sees and is aware of conflict) which in turn is a necessity for response (Lee et al., 2017). Finally, the drivers' trust in the automated system has been found to influence driver response to safety-relevant events during automation, but requires more work before the impact on the driver response process is fully understood (McDonald et al., 2019). Higher trust in automation have been found to result in increased response times and more collisions, compared to lower trust levels (Körber, Baseler & Bengler, 2018; Payre, Cestac & Delhomme, 2016).

1.4 Human (cognitive) mechanisms and frameworks for explaining automation effects

As discussed in previous chapters, when vehicle automation fails, drivers may have troubles performing as well as they do in manual driving (i.e. we observe automation effects during assisted driving and aftereffects after unsupervised automation). This effect of automation on the driver response process needs to be well understood, as it may have serious consequences such as crashes. Recall that safe and effective driving depends largely on the human ability to perceive detect and gather information about the environment, comprehend this information and act appropriately.

Specifically, visual attention has been pointed out as one of the most significant aspects of safe manual routine-driving and event-response (Macadam, 2003; Victor et al., 2015). Therefore, the knowledge of reduced monitoring because of increased automation as for example pointed out by Morando (2019), is an obvious safety concern. Unsupervised automation may even allow a reduced visual attention from the system design, since drivers are free to engage in non-driving-related activities and consequently do not need to pay attention to the forward road. However, during assisted driving there is a current challenge in how to secure appropriate visual attention levels, due to the knowledge of humans being poor at supervising (Warm, Parasuraman, & Matthews, 2008). Recall that appropriate attention levels (i.e. drivers that look on the forward road) are needed because during assisted driving the drivers are responsible for detecting and acting in safety-relevant events which may not be prompted by the system.

One way to look at attention is in terms of a filter that guides or selects the information that will be part of higher-level human cognition (Lee et al., 2017). The environmental information we attend to depends on: the salience of the information (e.g. a very loud alarm), the effort needed (e.g. do we need to turn our head around), the expectancy of valuable information (e.g. a person often crosses the road at a certain place) and the value of (or cost of not) attending to a specific stimuli (e.g. if not looking at the forward road we may crash). However, just because humans look at something, it does not mean they attend to or perceive it. For example, the phenomenon known as *attentional*

blindness means that a fully visible object is missed because attention was devoted somewhere else (Lee et al., 2017).

There are many proposed cognitive mechanisms (with links to visual attention) used to explain why humans perform worse after automation compared to after manual driving. For example a delayed response in safety-relevant events after automation may stem from reduced monitoring related to *overtrust* in the automated system (Parasuraman & Riley, 1997), a reduced *situation awareness* (Endsley, 1995, 2015), or drivers that are *out-of-the-loop* (Endsley & Kiris, 1995). The construct of situation awareness (SA) can shortly be described as: "internal conceptualization of the current situation" and is formed on following three levels: (a) drivers perceive elements of the environment ("Which information do I need?"), (b) they comprehend the meaning of these elements ("What does this mean to me?"), and (c) they predict their near-future status ("What do I think will happen next?") (Bolstad et al., 2010). The out-of-the-loop construct is attributed to the possible loss of manual operational skills and a reduced situation awareness that may stem from the introduction of automation (Endsley & Kiris, 1995). Merat et al. (2019) proposed a definition of the *out-of-the-loop* in the context of vehicle automation. Within this context, the *out-of-the-loop* phenomena is defined as "not in physical control of the vehicle and not monitoring the driving situation, OR in physical control of the vehicle but not monitoring the driving situation". For completeness, *in-the-loop* is defined as: "in physical control of the vehicle and monitoring the driving situation" and *on-the-loop* "not in physical control of the vehicle, but monitoring the driving situation" (Merat et al., 2019). Other examples of proposed cognitive mechanisms associated with automation effects are, drivers that are mentally underloaded (Young & Stanton, 2002), drivers that have insufficient mental models of the automated system (Victor et al., 2018) or the disruption of the perceptual-motor loop as a consequence of automation (Mole et al., 2019).

In addition to previously mentioned human mechanisms, several frameworks or models have been developed to understand how these human (cognitive) mechanisms interact with action to produce responses. For example, the *information processing model of cognition* represents the human information-processing as consisting of four stages between which information gets transformed (Lee et al., 2017; Wickens, 2002). That is, (a) we sense the environment, (b) we perceive it's meaning based on what we sensed and prior knowledge (referred to as bottom-up and top-down processing), (c) we manipulate the information in our brain either through central processing (e.g. selects an action) or through transforming and remembering, and (d) we respond to the information (e.g. executes the action). All these stages are dependent on a limited pool of attentional resources: if one stage requires a lot of attention another step may be degraded. This division of information processing into different steps has also been proposed in other models (e.g. Endsley, 1995).

Recently, another framework towards understanding the human cognition and action process in driving has been proposed. This new framework is called *predictive processing,* or short PP (Clark, 2013; Engström et al., 2018). It differs from the traditional information-processing assumption of a feed-forward stream of information from sensation to action. According to the PP framework, the brain continuously predicts sensory input from the external environment (e.g. looming – the visual expansion of an approaching lead vehicle on the retina) and minimizes deviations between predicted and perceived sensory inputs, through action (e.g. braking, steering) or by updating the prediction (Clark, 2013; Engström et al., 2018). These predictions

are generated by a hierarchical generative model, which is embodied in the brain and develops with experience.

Engström et al. (2018) also proposed a way of explaining the out-of-the-loop phenomena in automated driving through the concept of *active inference* and the different levels of the driving task (Michon, 1985). In the context of manual driving, the continuous minimization of prediction errors by either updating predictions or acting (steering, braking) is referred to as active inference. Active inference may take place on both the operational, tactical and strategic levels of the driving task. Driving manually is to be part of active inference on all three levels (i.e. to be in the loop on all three levels). However, when automation is introduced, active inference may not take place on some of these levels (e.g. the operational or tactical levels), and consequently the driver may be out-of-the-loop on some of these levels. In assisted driving, when the vehicle performs longitudinal and lateral control, but the driver needs to monitor the environment, the driver is assumed to be engaged in perceptual inference rather than active inference on the operational level. This means that the driver may still make predictions on looming by observing a lead vehicle in front but is not involved in physical vehicle control on the operational level. Thus, the driver (if still making predictions on looming) may still be in the loop on the operational level, even if the type of inference (active vs. passive) is different from manual driving. However, if the driver is only monitoring the environment without making predictions about looming, s/he is assumed to be out of the operational loop but may still be involved in making inferences on the tactical and strategic loop (i.e. in-the-loop on these levels). During unsupervised automation when a driver disengages fully from the driving task and is engaged in a non-driving related task, the driver may be out of both the tactical and operational control loops.

1.5 Research needs

To achieve safe vehicle automation, the current understanding of the scenarios which are associated with automation effects needs to be enriched. There is a need to understand if the automation effects observed in driving simulators also exist in more realistic settings (e.g. on test track with a real vehicle). Specifically, the two automation types *assisted driving* and *unsupervised automation* requires more (and individual) attention. This is because: (a) assisted driving systems in on-market vehicles have been involved in crashes in real traffic, (b) the factors (contributing or mitigating) that influence these crashes are not fully understood, (c) factors influencing driver response in assisted driving may not be the same for unsupervised automation (due to different driver roles and system features e.g. presence of prompts, for the two automation types), and (d) future automated vehicles will likely be equipped with versions of these two types of automation. In fact, for both automation types, detailed investigations of the complete driver response process in safety-relevant events are a necessity. The reason is that automation effects may: (a) only occur some time after automation deactivation and would then be missed if only considering for example the TOT, (b) be influenced by the time needed for drivers to perform the actions in the response preparation phase (i.e. the preparation-action-time consequence as introduced in Section 1.3.3) or (c) be caused by some human mechanisms (cognitive or non-cognitive) that can be understood only by detailed investigations of the complete driver response process (e.g. a mechanism that may only influence the driver steering control after automation has been deactivated; Mole et al., 2019). Finally, a hands-on-wheel requirement and automation duration may influence driver response, but current evidence is either lacking or points in different directions.

2 Objectives

The overall aim of this PhD project is to contribute to the development of *safe* vehicle automation, so that it can be objectively measurable. To obtain objectively measurable safe vehicle automation, there is a need to understand the factors that influence the existence of automation effects, and the components of these effects, through investigations of the driver response process. Importantly, there is also a need to advance the ecological validity of the current understanding of automation effects that mainly stems from driving simulator studies, by using data collected in more realistic settings. Specifically, this PhD project contributes to this overall aim, through empirical studies and analyses of the driver response process in safety-relevant events in automation from test-track and naturalistic data. To achieve this aim, five objectives were specified, out of which the first two are addressed in this licentiate thesis:

1. To investigate the actions in the response preparation phase of the driver response process in a safety-relevant event in assisted driving when no prompt is present (ADS-non-prompted), and specifically the influence of a hands-on-wheel requirement, on test track.

2. To investigate the driver response process in a safety-relevant event in unsupervised automation when a prompt is present (ADS-prompted), and specifically the influence of automation duration and timings for the conflict onset and the prompt, on test track.

The continued work in this PhD project will likely focus on:

- Investigations of the driver response process in the same safety-relevant event as in *objective 1* after a period unsupervised automation when a prompt is present (ADS-prompted), and specifically the influence of the duration between the timings for the conflict onset and the prompt, on test track.

- Safety analyses of the driver response process in unsupervised automation when a prompt is present (ADS-prompted) before the driver needs to resume manual control from automation, in a naturalistic setting.

- Modelling of the driver response process for different automation types (assisted driving and unsupervised automation) and prompts (TOR vs. ADS-non-prompted) in a safety-relevant event.

3 Methods

This Chapter gives an overview of the methods, settings, and tools that can be used to study driver behaviour in automated driving, and specifically the combination of methods, settings, and tools that were used to study the driver response process in the two papers included in this thesis (Paper I and II). In addition, this Chapter introduces experimental protocols, driving measures, and statistical methods that can be used to assess the driver response process to safety-relevant events.

3.1 Methods for studying driver behaviour in automated driving and participant selection

Methods for studying driver behaviour typically require a compromise between the *experimental control* and *realism* (McLaughlin, Hankey, & Dingus, 2009). On one side, we find the *controlled studies* in which two or more independent variables are typically manipulated, whereas the remaining factors that may influence the measured dependent variable/s are kept fixed. Keeping some factors fixed may create an artificial environment which may influence how humans behave, and thus the extent to which results in such a setting generalizes to real-world settings remains unknown. On the other side of the spectrum are the naturalistic studies in which drivers are observed (i.e. their normal car is instrumented with sensors such as cameras) when they drive as they normally would. In these naturalistic studies, the degree of realism is high, whereas understanding relationships between different factors is a challenge because factors cannot be controlled.

Examples of test environments for studying driving behaviour include driving simulators, test tracks, and public roads (McLaughin et al., 2009). A driving simulator enables experiments with a high degree of control and the possibility to include critical safety-relevant events that may result in crashes with little ethical concern (Fisher, Rizzo, Caird, & Lee, 2011). Even if driving simulators offer a safe and convenient way of measuring driving performance, the results of driving simulators may not generalize to driving in the real world. Specifically, driving simulators have not been found to be able to reproduce absolute values, and sometimes also not absolute differences between conditions (i.e. they lack *absolute validity*). However, driving simulators have been able to produce differences in the same direction (e.g. speed reduction), when compared to real-road testing (i.e. they have *relative validity*) (Fisher et al., 2011). Thus, when absolute values are required, tests in realistic environments such as as test-track experiments or public-road testing is generally necessary. Much of what is known today about the driver response process during assisted driving and unsupervised automation stems from driving simulator experiments.

Test-track experiments offers a higher degree of realism than driving simulators (e.g. real visual and kinematic cues), but lack some of the control (McLaughin et al., 2009). In addition, test-track experiments can be more controlled than on-road studies, since safety-relevant events can be included with higher repeatability and safety compared to test on public road (even if some ethical restrictions apply to how critical these events can be). Test-track experiments also lack some degree of realism since an experimental leader is often present in the car. The data used in the two papers included in this thesis (Paper I and II) was collected in two test-track experiments which included two safety-relevant events, which would not have been possible in a public road study and very artificial in a driving simulator. Experiments that do not require the safety and the control offered by experiments performed in driving simulators and on test track, can be performed on public roads. These studies offer the opportunity to test how drivers

may interact with systems in an environment with surrounding traffic and where normal driving hazards are present.

3.1.1 Participant selection

In order to understand how drivers respond in a safety-relevant event during assisted driving and after unsupervised automation, human subjects to be studied are needed. However, to make statements about a population based on a studied sample (i.e. a selection of participants from a population) it is important to consider who the selected participants are (e.g. age, gender, education) and where these participants come from (e.g. Sweden, a specific region in Sweden etc.; University of Michigan, 2018). The participants included in Paper I and II were all Volvo Car employees working in Gothenburg, Sweden. To minimize biases, the participants had no work duties associated with the development of automated driving, did not work as test drivers, and had not been part of similar studies before. All included participants had driven at least 5000 km during the year prior to the study. Both samples were to the extent possible both age- and gender-balanced.

3.2 Wizard-of-Oz vehicle

While automated functions can be easily simulated in a driving simulator, test-track and public road testing requires a real vehicle with a reliable automated system. In lack of reliable unsupervised automated vehicles, there is a need to find other ways of investigating human collaboration with vehicle automation. One such approach is the *Wizard-of-Oz* technique. Generally within the field of human-computer interaction, a wizard-of-oz experiment (previously *Oz paradigm*) is an experiment in which participants think that they interact with an automated system, but in reality the automation is simulated by a human who is often partly or fully hidden (Kelley, 2018). Practically, the *Wizard-of-Oz* technique can be implemented in a real vehicle to study humans interact with vehicle systems, for example by enabling vehicle control to be executed from somewhere else than the drivers' seat (Habibovic, Andersson, Nilsson, Lundgren, & Nilsson, 2016). These W*izard-of-Oz vehicle*s can be used for experiments on test track and on public roads.

The two test-track experiments (Paper I and II) included a Wizard-of-Oz vehicle. This vehicle was a Volvo passenger car which was rebuilt to include a steering wheel and pedals positioned in front of the middle backseat. Both the steering wheel and the pedals were hidden from the participant in the driver seat. This setup enabled for a test driver (the *Wizard*) to control the vehicle from the backseat. The wizard-of-Oz vehicle used in the experiments for Paper I and II were equipped with three cameras that recorded the forward road, the driver's face and the driver's upper side body. In addition, typical vehicle signals (e.g. vehicle speed and acceleration), GPS signals, and signals that capture when the TOR was issued and when the driver had deactivated automation, were collected.

3.3 Experimental protocols to investigate the driver response process during assisted driving and after unsupervised automation

Experiments performed to be able to assess the driver response process after a period of assisted driving or unsupervised automation can be designed in different ways. Typically, the experimental protocols include a period of driving with an assisted-driving- or an unsupervised-automation-system engaged, which leads up to a safety-relevant event which require the drivers to start to drive manually to avoid a crash (Gold

et al., 2013; Happee et al, 2017; Louw et al., 2015; McDonald et al., 2019). The safety-relevant event can for example consist of a braking lead vehicle combined with a silent ACC failure (Bianchi Piccinini et al., 2019), a cut-in or cut-out scenario (Larsson et al., 2014; Victor et al., 2018), or a conflict object (e.g. a broken-down vehicle) in lane that drivers need to avoid by steering or braking (Gold et al., 2013; Louw et al., 2015).

Specifically, the experimental protocols used in Paper I and Paper II included two safety-relevant events which included an object (the conflict object) that required for the drivers to act. In Paper I, this event was a cut-out scenario with either a garbage bag or a stationary vehicle as the conflict object positioned in the lane (Scenario 1 in Figure 1 shows this setup). Note that the data used in Paper I was a subset of the complete dataset used in the work by Victor et al. (2018). In fact, Paper I aimed to enrich the understanding of conflict outcome (crashing or avoiding the conflict object) observed in Victor et al. (2018) by performing detailed analyses of the driver actions in the response preparation phase prior to the crash/avoid point. In Paper II, the event was a road-work zone built up of cones (the road-work zone will be referred to as the conflict object in this case) which was revealed by a lead vehicle that changed lane (Scenario 2 in Figure 1 shows this setup). In Paper I, the need for drivers to start to drive manually to handle the cut-out scenario was ADS-non-prompted since the drivers needed to detect the upcoming conflict object (a garbage bag or a stationary balloon vehicle) and act not to crash with it. In Paper II, on the other hand, the need to start driving manually before the upcoming road-work zone was ADS-prompted since a TOR was issued 5-6 seconds before the lead vehicle changed and the road-work zone became visible to the driver.

3.3.1 The preparation-action time consequence

In fact, the type of setup used in Paper II differed from previous studies (Gold et al., 2013; Happee et al., 2017; Louw et al., 2015), since the TOR was issued before the conflict onset (Figure 2, bottom row). This setup enabled for the drivers to complete the actions within the response preparation phase (e.g. put their hands on the wheel) *before* the lead vehicle changed lane. Paper II also included a manual baseline. A manual baseline (Figure 2, top row) facilitates understanding of whether the observed driver behaviour (e.g. crashing) is due to the period of assisted driving or unsupervised automation, or simply happened because the situation was outside human ability. For example, a very critical safety-relevant event may not be avoided because of the fact that humans show time delays in reacting to stimuli (e.g. a suddenly appearing conflict object; Macadam, 2003).

When the driver response process after unsupervised automation is compared to the driver response in manual driving, the timings for the *prompt (TOR)* relative to the *conflict onset* is important. This is because of the preparation-action time consequence which was introduced in Chapter 1, Section 1.3.2. Recall that this consequence refers to the automation effects that stem from the time needed for drivers to prepare for action after automation. When the timing for the prompt (TOR) and the timing for the conflict onset align (see Figure 2, middle row), the drivers in automation need to complete the actions within the response preparation phase before the driver action start, whereas the drivers in the manual condition can act directly assuming these drivers are fully engaged in the driving task. Thus, previous studies that have used the setup illustrated in Figure 2 (middle row) may have been biased: they may have observed automation aftereffects (e.g. delayed response, a degraded manual driving performance, or crashing, after automation was deactivated) that were simply a consequence of the preparation-action time. In other words, this preparation-action time consequence means that the response

preparation introduces a time delay which is typically not needed when drivers are already in manual driving mode.

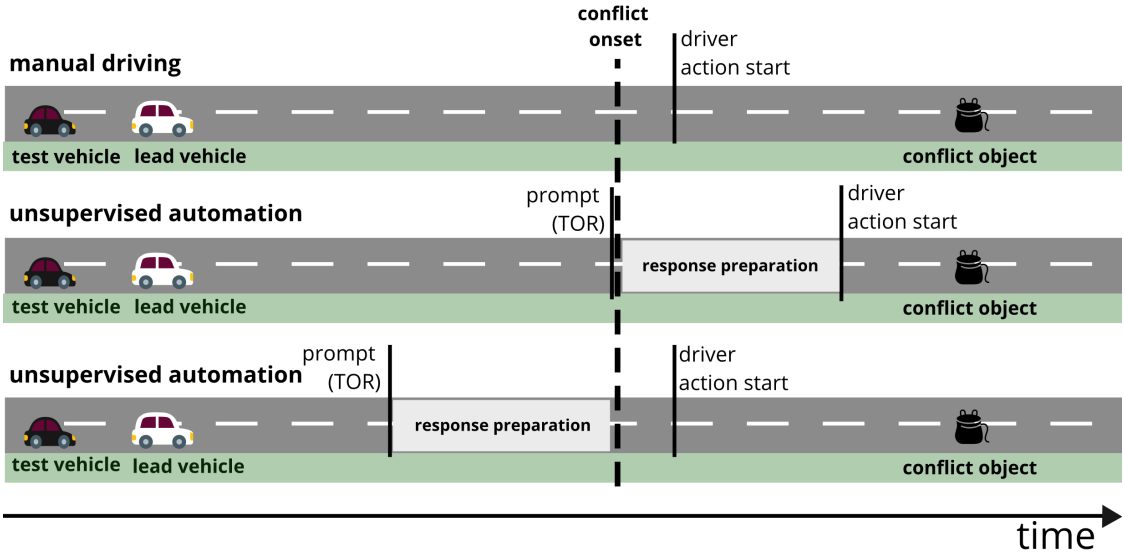

Figure 2 - A representation of how the timings of the prompt (TOR) and conflict onset matter for when drivers can start acting to avoid a conflict object as part of a safety-relevant event. Each row represents the same safety-relevant event which is encountered in manual driving (top tow) and unsupervised automation driving (middle and bottom row). The only difference between the middle row and the bottom row is when the prompt (TOR) is issued in relation to the conflict onset.

3.4 Data processing and analysis

The driver response process can be used to assess the driver response to a safety-relevant event through: (a) decomposing reaction times into its consisting time-components, and (b) analyses of the intervention profile (Lee et al., 2002). For example, Lee et al. (2002) analysed brake responses in a rear-end collision event through decomposing the brake reaction time (the time from a warning to the point when the driver began to decelerate) into four components (e.g. accelerator release reaction time, accelerator-to-brake transition time etc.) and used the mean and maximum brake accelerations as metrics for assessing the brake profile. The driver response process can also be used to assess drivers' response to a safety-relevant event after a period of assisted driving or unsupervised automation (Morando, Victor, Bengler, & Dozza, 2020, Gold et al., 2013). For example, Gold et al. (2013) decomposed the TOT into smaller components (e.g. reaction times for positioning hands on the steering wheel, redirecting eyes to the forward path). Further, Gold et al. (2013) assessed the manual intervention performance with for example lateral vehicle position trajectories in the safety-relevant event and with the *utilization of the acceleration potential* (i.e. the square-root of the sum of the squared maximum longitudinal and lateral accelerations).

In order to capture the driver actions that make up the driver response process, video data is needed. That is, driving simulators and test cars can be equipped with video cameras that record the driver from different angles (e.g. a video camera positioned to capture the drivers face). Then, in order to extract time points for the driver actions of interest for the driver response process, manual video annotation is usually performed. That is, one or several people observe the video and markdown time stamps when a certain action appears (e.g. driver has at least one hand on the steering wheel). These time stamps can then be extracted and combined into a dataset which can be used to

analyse the response process. For analyses of the intervention and driving performance, a combination of *discrete metrics* (e.g. maximum lateral acceleration, minimum steering wheel angle) and *continuous metrics* (e.g. full vehicle speed signals) can be used. The risk of only using discrete metrics is to miss important information to be able to assess the safety of the driver response process. For example, a significantly different lane position directly after drivers have deactivated automation may be a sign of that the drivers repositioned the hands on the steering wheel and applied some torque, and nothing necessarily safety critical. However, while discrete metrics can easily be included in statistical analyses to understand effects and effects sizes, continuous metrics may require more advanced methods.

In this thesis, the driver response process consists of two components: the *response preparation* phase and the *manual intervention and stabilization* phase (see Figure 3). The response preparation phase includes all actions that are performed up to the *driver action start* and the manual intervention and stabilization phase consists of the manual driving performance that follows after the driver has started driving manually.

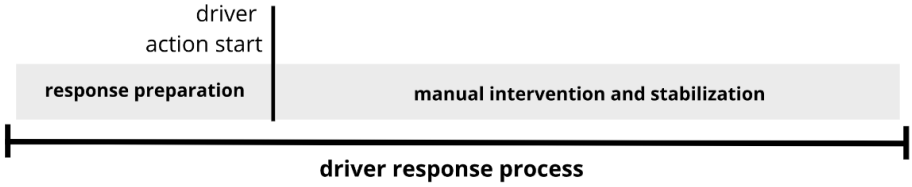

Figure 3 - A representation of the driver response process and its two components (the response preparation and the manual intervention and stabilization). The response preparation includes driver actions needed to be performed before the driver can start to drive manually, whereas the manual intervention and stabilization includes the manual driving performance that follows the driver action start.

In order to analyse the driver response process in Paper I and Paper II, manual video annotation was performed to extract time points for driver actions performed in the safety-relevant events. Examples of these times points are: when a driver showed a surprise (SRT), put hands on wheel (HOW), started steering to avoid the conflict object (DS), redirected the eyes from the mounted tablet to the human-machine interface (EY HMI), started to look on the forward path (EY FWD) and when the drivers deactivated automation (AD DEACT; examples of these time points are shown in Figure 1). These time points (or metrics) were mainly used to assess the response preparation phase. In Paper I, the driving performance in the manual intervention and stabilization phase were simply assessed with the conflict outcome (i.e. if the drivers crashed with or avoided the conflict object). In Paper II, the analyses of the driving performance in the manual intervention and stabilization phase was extended to include more detailed analyses of the driver intervention performance (e.g. trajectories for vehicle speed and accelerations when drivers manoeuvred the road-work zone combined with metrics for discrete maximum speed and maximum lateral and longitudinal accelerations).

To mathematically assess how meaningful an observed difference between two or more metrics (as part of the driver response process) are, statistical methods were used in both Paper I and II. In fact, two different types of statistical methods were used, namely frequentist methods (Paper I) and Bayesian methods (Paper II). Frequentist methods are by far the most commonly used within the human factors in automation literature, why it was also used in Paper I. However, frequentist methods (especially the so-called null hypothesis significance testing, NHST) have the disadvantage of inducing a black-and-white thinking were effects either exists or do not exist (Kruschke & Liddell, 2018). The real world is often more nuanced, which Bayesian methods are better at capturing.

One reason is that the output of a Bayesian analysis is a distribution of a parameter (e.g. the mean) together with the uncertainty of this parameter value. Thus, the output includes both possible magnitudes as well as probabilities of these magnitudes. This is more informative than the information given in NHTS (i.e. if a *p*-value is rejected or accepted without explicit information about parameter magnitudes). Information about actual magnitudes enables researchers to assess the relevance of results in other contexts. For example, the difference between two parameters (e.g. difference in vehicle speed) may be proven statistically significant with a NHTS, but the actual difference in magnitude (e.g. 0.1 m/s) may be very small and not meaningful in a specific context, and Bayesian methods enables the reader to make this assessment. In Paper II, Bayesian methods were used to estimate differences between metrics across conditions (e.g. maximum vehicle speed after automation and after manual driving) together with the uncertainty of this difference.

4 Results

This Licentiate thesis includes two journal papers, reported in Table 1. Paper I and Paper II are submitted to two of the leading journals within the human factors in automation field. This Chapter presents summaries of these two papers.

Table 1: Appended papers

Papers	
Paper I	**Driver conflict response during supervised automation: do hands on wheel matter?** **Pipkorn, L.**, Victor, T., Dozza, M., Tivesten, E. (2020). Driver conflict response during supervised automation: do hands on wheel matter? *Transportation Research Part F: Traffic Psychology and Behaviour.* Submitted.
Paper II	**Automation aftereffects: the influence of automation duration, test track and timings** **Pipkorn, L.**, Victor, T., Dozza, M., Tivesten, E. (2020). Automation aftereffects: the influence of automation duration, test track and timings. *IEEE Transactions on Intelligent Transportation Systems.* Submitted.

Paper I. Driver conflict response during supervised automation: do hands on wheel matter?

Introduction Understanding how to secure appropriate driver response during supervised automation is an important step in achieving safe automation. However, in-depth knowledge regarding the mechanisms affecting driver response process is lacking.

Objective The first aim of this study was to investigate how driver conflict response in supervised automation differ for drivers that crash and drivers that avoid a conflict object in a critical event. The second aim was to understand the influence of three specific factors in the driver response process: a hands-on-wheel requirement (with vs. without), the conflict object type (garbage bag vs. stationary vehicle), and the driver trust level (high vs. low).

Method Seventy-six participants supervised automation for 30 minutes on a test track, before encountering a conflict event. In the conflict event, the participants needed to avoid a conflict object which was revealed by a lead-vehicle cut-out. The driver conflict response was assessed through the response process: timepoints for driver surprise reaction, hands-on-wheel, drivers steering, and driver braking.

Results Crashers generally responded later in all actions of the response process compared to non-crashers. A hands-on-wheel requirement did not influence driver conflict response: the drivers with and without hands on the wheel started steering to pass the conflict object at similar times. High-trust drivers generally responded later, than the low-trust drivers or not at all, and only high-trust drivers crashed. The larger stationary vehicle triggered an earlier surprise reaction compared to the garbage bag, while hands-on-wheel and steering response were similar for the two conflict object types.

Discussion A hands-on-wheel requirement may not prevent drivers to respond late or crash, when drivers have supervised automation for some duration. To what extent this result generalises to other types of conflicts (e.g. sideswipes, lane exits) is currently unknown. In addition, further research is also needed to understand to what extent a hands-on-wheel requirement that also requires a certain amount of torque input would give other results.

Paper II. Automation aftereffects: the influence of automation duration, test track and timings

Introduction Automation aftereffects (i.e. degraded manual driving performance, delayed responses, and more aggressive avoidance maneuvers) have been found in driving simulator studies. In addition, longer automation duration seems to result in more severe aftereffects, compared to shorter duration. The extent to which these findings generalize to real-world driving is currently unknown.

Objective The aim of this study was to examine the effect of automation exposure and its duration on the *driver take-over response* and *driving performance* in a simulated road-work zone on test track. In addition, by comparing the present study's results with previous driving simulator studies, this study also aimed at better understand the influence of different factors (e.g. test environment, experimental protocols) on the automation aftereffects.

Method Seventeen participants followed a lead vehicle on a test track. They encountered the road-work zone three times: while driving manually, and after a short and a long duration of automation. The take-over request was issued 5-6 s before the lead vehicle performed a cut-out and revealed the road-work zone.

Results All drivers managed to resume manual control in response to the take-over request, and then manoeuvre through the cone zone with a similar driving performance as in manual driving, and without colliding with any cones. The effects of automation (i.e. automation vs. manual) were greater than the effect of automation duration, but in contrast to previous driving simulator studies, the observed effects were minor.

Discussion The automation aftereffects observed in the present study were not as large as previously found in driving simulator studies. To what extent this observation is due to the use of different test environments (driving simulator vs. test track) or different experimental protocols is unknown. However, independent of test environment, in the search for automation aftereffects it is important to consider the influence of the driver take-over response, on the observed aftereffects. That is, more work is needed to disentangle the aftereffects that is merely a result of a longer driver take-over response process, and the aftereffects that may be caused by some other human (cognitive) mechanism (e.g. situation awareness, out-of-the-loop, less calibrated sensorimotor control).

5 Discussion

5.1 Driver response process during assisted driving

Paper I shows that automation effects (e.g. delayed response, degraded driving performance, or crashing involvement) may exist when drivers need to act in a lead-vehicle cut-out scenario during assisted driving when no prompt is present. In fact, despite having eyes on the threat, some drivers responded later in all actions of the response preparation phase of the driver response process (i.e. surprise reaction, hands on wheel, driver steering start), whereas some only showed a surprise reaction without putting hands on the wheel or attempting to steer. This delay in—or lack of—response observed in Paper I is in line with previous research on driver conflict response to ADS-non-prompted events during driving with different degrees of assisted driving (such as ACC or ACC with Automated steering; Larsson et al., 2014; Bianchi Piccinini et al., 2019; Strand et al., 2014; Young & Stanton, 2007).

5.1.1 Advantages of the response preparation phase and the influence of a hands-on-wheel requirement

Through its detailed analyses of the actions within the response preparation phase, Paper I enabled an enhanced understanding of the way these actions were associated with the conflict outcome as previously identified in Victor et al. (2018). In addition, these detailed analyses provided insights into how three individual factors (i.e. a hands-on-wheel (HoW) requirement, driver trust, and conflict object type) influenced the actions and consequently the response process. For example, the findings suggest that: (a) some drivers crashed without putting their hands on the steering wheel, whereas others who also crashed mainly showed a delay in the hands-on-wheel- or steering response, (b) high-trust drivers generally put hands on wheel and started steering later than low-trust drivers, and (c) a larger conflict object only influenced the timing of the surprise reaction, but not the hands-on-wheel or steering response times. This detailed analysis of the response preparation phase is in contrast with previous research on driver conflict response to ADS-non-prompted events during assisted driving that mainly focus on a single response time such as the brake response time (Larsson et al., 2014; Bianchi Piccinini et al., 2019; Young & Stanton, 2007).

One of the main findings in Paper I was that a HoW requirement and supervision reminders during assisted driving did not help the drivers to avoid crashing nor did it elicit an earlier steering response. This finding is in line with previous studies that issued a TOR prior to a safety-relevant event (Naujoks et al., 2015, 2017), but contrasts to Llaneras, Cannon, & Green (2017) who used a HoW-requirement as an *escalation of consequence* when drivers ignored visual attention reminders in a (silent failure) lane-drift event. Thus, a HoW requirement may still be beneficial in other types of conflicts (e.g. lateral lane drifts or incorrect system steering). The extent to which a modified HoW requirement in the present study would be able to mitigate crashing remains unknown. For example, a HoW requirement inspired by the work of Llaneras et al. (2017) could include different types of required physical (hands-on-wheel) involvement which would depend on the drivers' engagement in supervising the assisted driving system. For example, when a driver is examined as insufficiently engaged (and does not change behaviour in response to a requirement or reminder) s/he may first be required to rest the hands on the steering wheel, then resist the system-initiated torque and finally may need to actively provide steering torque. These different types of hands-on driving represent different degrees of *physical control*, and thus could potentially be considered different types of being in-the-loop according to Merat et al.

(2019). To conclude, more work is needed to understand the physical involvement necessary to facilitate safe and appropriate response to a safety-relevant event during assisted driving when no prompt is present.

5.1.2 Factors explaining delayed response and crashing during assisted driving

Victor et al. (2018) concluded that the drivers crashed because of an *automation expectation mismatch* i.e. drivers crashed because they expected automation to avoid the object in the cut-out scenario, despite having eyes on the threat. This conclusion was based on interviews after the drive. In fact, the drivers who crashed reported that they expected automation to act in the conflict, whereas the non-crashing drivers reported that they either were uncertain or did not expect automation to act (Gustavsson et al., 2018; Victor et al., 2018). Paper I confirmed that all drivers showed high levels of visual attention to the forward path in the safety-relevant event. In addition, Paper I enriched this understanding further by also finding that all drivers except one showed a surprise reaction in the safety-relevant event. This facial sign of surprise may be an indication of that the drivers were aware of the conflict object. Thus, Paper I confirms that drivers did not seem to crash merely because they did not detect or were not aware of the conflict (e.g. due to low levels of visual attention).

In addition, Paper I found that drivers that reported high trust in automation, responded later than the drivers that reported low trust in automation. The predictive processing (PP) framework (Clark, 2013) was introduced in Paper I as a possible explanation for the difference in conflict outcomes and response due to the reported trust in automation. Explaining the results within the PP framework was novel: despite the recent advances of the PP within cognitive neuroscience, this cognitive framework is rarely used to explain results in the literature on human factors in automation. Simply put, the delayed response and consequent crashing reported in Paper I may arise from the crashers and the non-crashers having different understandings of the assisted driving systems capabilities (in the PP framework, this mismatch can be explained as crashes and non-crashes having different hierarchical generative models). Further, the difference in response for the high-trust drivers who crashed and the ones that avoided crashing may arise from these drivers being differently involved in the driving task on the operational level (Michon, 1985). In the PP framework, this can be explained as the high-trust drivers who avoided crashing was involved in perceptual inference (they were making predictions on looming) on the operational level, whereas the crashing high-trust drivers were not engaged in any inference on the operational level. Other frameworks or conceptual models may also be useful to explain the results, but these were not considered within the scope of this thesis.

5.2 Driver response process after unsupervised automation

The finding of automation effects in Paper I, that combined assisted driving, no prompt, and an unexpected critical event, motivated the work in Paper II which aimed to understand if automation effects would also be present in another scenario, namely a safety-relevant event that was assumed to be easier for the drivers to handle. This was realized in Paper II by combining unsupervised automation that could issue a prompt (TOR) prior to a safety-relevant event that was more expected than the one in Paper I because the drivers had already encountered the same event twice before.

Interestingly, despite the increased automation and the fact that the drivers in Paper II were out-of-the-loop (the drivers were not involved in physical control and did not

monitor the driving environment; Merat et al. 2019), only minor automation aftereffects (i.e. automation effects specifically occurring after automation has been deactivated) were observed. That is, all drivers after automation, started their steering manoeuvre earlier or at similar timings, and showed a similar driving performance in the safety-relevant event compared to the manual pre-event condition. The extent to which the difference in automation effects between Paper I and II depend on different types of safety-relevant events (i.e. both criticality and expectancy), type of automation, and presence/absence of a prompt prior to the safety-relevant event should be further investigated. Whereas, increased automation has been found to result in poorer performance (Onnasch et al., 2014), a less critical and more expected safety-relevant event is likely to result in improved driving performance compared to a more critical and less expected event. In addition, prompts that call for drivers' attention have been found to decrease response times (Lee et al., 2002), and the presence of a TOR in Paper II, in contrast to Paper I, could therefore partly explain the appropriate response observed in Paper II.

It also seems that, if a prompt (TOR) is triggered at a sufficient time before conflict onset, the drivers are able to re-engage in the driving task on both the operational and tactical levels (Michon, 1985) and start making predictions on looming (as assumed within the PP framework) in a timely manner to avoid crashing in the safety-relevant event. The fact that the drivers responded in the event (i.e. the drivers acted) is evidence for that these drivers must have generated prediction errors to be acted upon, according to the PP framework (Clark et al, 2013; Engström et al. 2018). Assuming that the drivers were not making predictions on looming while in automated mode (i.e. the drivers looked on the game they played and not on path), these drivers must have then started to make predictions on looming in order to generate prediction errors to act upon.

The finding of minor automation aftereffects in Paper II, differ from the findings of significant aftereffects in previous studies in driving simulators on unsupervised automation and prompted safety-relevant events (Gold et al., 2013; Louw et al., 2015; Happee et al., 2017). One likely reason explaining these different findings is that the drivers in Paper II were able to resume manual control before the conflict onset (Figure 2, bottom row), whereas previous studies prompted the drivers at the conflict onset (Figure 2, middle row). Thus, Paper II indicated the importance of the *preparation-action-time consequence* as a possible explanation for observed automation aftereffects. This means that automation aftereffects in terms of delayed response, a degraded driving performance and crashing, after a period of unsupervised automation, may stem from the time it takes for drivers to prepare for action after automation (e.g. reposition hands to the steering wheel). However, automation aftereffects may also stem from any of the cognitive mechanisms explained in Chapter 1, Section 1.4 (e.g. mental underload, reduced situation awareness). In other words, when automation prompts drivers to transition at the conflict onset, delayed response (and potentially crashing) can stem from drivers being less aware or mentally underloaded (cognitive mechanism), but is also affected by the time needed for drivers to become ready-to-act after automation (position hands to the steering wheel, deactivate automation etc.). Thus, careful experimental design, which considers the consequences of the preparation-action-time is needed before automation aftereffects may be attributed to some cognitive mechanism (e.g. reduced situation awareness). In addition, the preparation-action-time needs to be considered to understand the influence of automation on the steering behaviour (e.g. as hypothesized by Mole et al., 2019). The reason is that an observed degraded steering performance after automation may stem from less calibrated perceptual-motor control (Mole et al., 2019), but may also stem from the fact that

drivers after automation start steering closer to a conflict object and therefore are required by the situation to generate a more aggressive steering behaviour (as explained in Chapter 3, Section 3.3 and shown in Figure 2).

Finally, the results in Paper II suggest that automation duration has a minor influence on the driver response process. This finding contrasts to Bourrelly et al. (2019) and Jarosch & Bengler (2018) who found that a longer automated drive resulted in more severe aftereffects (e.g. a more degraded driving performance), but is in line with the findings of Feldhütter et al. (2017). In fact, the only observed negative effect of automation duration in on driver response, in Paper II, was that four drivers, after being exposed to automation for 14 minutes, had problems deactivating automation at the first attempt, while all drivers successfully deactivated automation at the first attempt after 4.5 minutes automation duration. This may be because a longer duration enables more time for drivers to forget the procedure for deactivating the automated system.

5.3 Contributions to safe vehicle automation

The overall aim of this PhD project is to contribute to the development of *safe* vehicle automation, so that it can be objectively measurable. This thesis demonstrates that both safe assisted driving and safe unsupervised automation can be achieved for most drivers in a realistic environment (on test track) because most of the drivers in Paper I and all drivers in Paper II were able to perform well in the safety-relevant event. However, since 28% of drivers still crashed in Paper I, more work is needed to understand how to prevent drivers from crashing in safety-relevant events encountered in assisted driving when no prompt is present. Paper II demonstrates that automation that prompts the driver to resume manual control may be highly efficient in helping drivers to respond appropriately in safety-relevant events that require driver actions to avoid crashing.

This thesis also demonstrates that safe vehicle automation can be objectively measured by investigating the driver response process in a safety-relevant event. Understanding the components of (e.g. the response preparation actions) and the factors (e.g. a hands-on-wheel requirement) that influence the driver response process is one important step towards developing safe vehicle automation. However, this information needs to be integrated in the design process in some way. Detailed understanding of the response process can be used to inform vehicle regulations for current and future assisted driving and unsupervised automation systems. For example, the findings in Paper I shows that requiring drivers to keep hands on wheel during assisted driving will not necessarily prompt earlier responses in longitudinal scenarios caused by automation limitation. However, as mentioned previously requiring drivers to keep hands on wheel may have other safety implications. The findings in Paper II suggest that drivers in unsupervised automation seem to be able to safely transition from automation to manual within 10 seconds as currently required in UNECE (2020). In addition, knowledge about the response process can be used to develop quantitative driver models that can be: (a) included in computational simulations for assessing the safety impact of different design choices (e.g. HMI, prompt design) or (b) inform driver monitoring systems. Since the driver response process differs for assisted driving when no prompt is present and unsupervised automation when a prompt is present, there may be reason to develop separate models for these fundamentally different automation types and driver roles. In addition, driver models for assisted driving when the need to resume manual driving is ADS-non-prompted need to capture the behaviour of drivers that crash due to an expectation mismatch, possibly through a model based on the PP framework that includes different generative models.

Finally, this thesis can also impact system design principles (e.g. HMI for transitions). Whereas, a difference of 9.5 minutes duration unsupervised automated driving does not seem to have any implications of drivers' ability to perform manual driving after a transitions, an improved HMI design can make the transition more intuitive and be less vulnerable for the tendency to forget how to deactivate the system after a long automation duration.

5.4 Limitations

The findings in Paper I and Paper II are based on experiments performed with a real vehicle on test track which provide a higher degree of realism than in driving simulators. However, a test-track study also has its limitations. The results may be influenced by the absence of real traffic, as well as the presence of a test leader and a safety driver in the vehicle. In addition, the conflict objects used in the two experiments lacked some realism for safety reasons. In experiment 1, the conflict object was a balloon car or a stuffed garbage bag, and in experiment 2 the conflict object was a simulated road-work zone built up of cones. However, it would not have been possible to perform a similar study which includes a safety-relevant event in a public road study with real traffic, even though the driver response process in non-critical scenarios could be investigated. In addition, the participants used in the two experiments were Volvo car employees in the Gothenburg area in Sweden who are not directly involved in product development of vehicle automation. Thus, the extent to which our results generalise to other populations remains unknown.

5.5 Future work

There is a need to understand in which way we can mitigate delayed response and crashing that may occur during assisted driving when no prompt is present. A first step could be to investigate if drivers would also crash in a lead vehicle cut-out scenario if this event is encountered during unsupervised automation when a prompt is present (since Paper II shows that drivers are able to response appropriately in an expected conflict). If drivers can respond appropriately in a safety-relevant event after having resumed manual control in response to a TOR, the expectation mismatch leading to crashes during assisted driving may be prevented in unsupervised automation when the system is capable of detecting the conflict and prompts the driver beforehand. As long as automation limitations contribute to crashes, however, we need to find ways to prevent these crashes from happening. For example, other types of designs that require the driver to put their hands on the steering wheel and apply torque occasionally may secure that the driver is sufficiently engaged in the driving task and respond to events when needed.

More work is also needed to understand the influencing factors behind automation aftereffects in unsupervised automation. Importantly, to guide the design of countermeasures, vehicle regulations, and driver monitoring, we need to understand if these automation aftereffects are simply a result of 1) the test environments (driving simulators vs. test track), 2) the time needed for drivers to become ready-to-act (i.e. to complete the response preparation), or 3) a cognitive explanation (e.g. drivers that do not make predictions on looming i.e. out-of-the-loop on an operational level according to the PP framework) or 4) a combination of the three. First, a driving simulator study and a test track study with the exact same setup could be performed in order to understand if the larger automation aftereffects observed in driving simulators are due to the different test environments, or if it is more related to the timings of prompt (TOR) and conflict onset. Then, following on the observations in Paper II, a study that further examines the influence of the TOR timing and the conflict onset timing on the driver

response process is needed. Specifically, since the take-over-time-budget influences the driver response process (Gold et al. 2013; McDonald et al., 2019), this should be varied.

Despite the advantages with test-track studies, there is still a need to understand how the driver response process may be influenced by the presence of real traffic with real threats. Thus, a study that investigates the driver response process in a more naturalistic setting (e.g. on public road) is needed. Finally, there is a need to develop driver models capturing the response process in safety-relevant events during assisted driving and unsupervised automation. The work by McDonald et al. (2019) represent a good start of the development of driver models for automated driving, by exploring how existing driver models (manual driving) may also be representative for predicting driver response in automation. As a step to continue that work, the analyses presented in this thesis could be turned into models representing the driver response to safety-relevant events during assisted driving and after unsupervised automation. As previously mentioned, such models could inform the design of driver monitoring systems and be used to assess the safety of certain design choices or countermeasures, in order to achieve safe vehicle automation.

6 Conclusions

This thesis has advanced the knowledge of human factors in automation through detailed analyses of the driver response process using test track data for both assisted driving and unsupervised automation, and by investigating the influence of different factors on the driver response process. Specifically, this thesis has advanced the understanding of automation effects and aftereffects, as well as the contributing factors behind, and the applicability of the predictive processing to explain, these effects.

Objective 1: To investigate the actions in the response preparation phase of the driver response process in a safety-relevant event in assisted driving when no prompt is present (ADS-non-prompted), and specifically the influence of a hands-on-wheel requirement, on test track.

Paper I showed that, all drivers may not be able to safely respond to a safety-relevant event (due to system limitation) during assisted driving, if they are not prompted by the automation. In fact, some drivers may show a delay in all actions (surprise reaction, hands on wheel, driver steering start) of the response process. Further, a hands-on-wheel requirement may not help the driver avoid crashing or elicit earlier steering response times in longitudinal conflicts that requires driver action. However, keeping at least one hand on the steering wheel may still prevent drivers from attending to dangerous secondary tasks and can potentially help drivers become aware of incorrect or insufficient system steering such as in lane drift events. More work is needed to understand if other types of a hands-on-wheel requirements (e.g. that requires drivers to actively provide steering input) can sufficiently keep the driver in-the-loop.

Objective 2: To investigate the driver response process in a safety-relevant event in unsupervised automation when a prompt is present (ADS-prompted), and specifically the influence of automation duration and timings for the conflict onset and the prompt, on test track.

Paper II showed that drivers may be able to safely resume manual driving when prompted by unsupervised automation to take over and respond appropriately in a subsequent safety-relevant event when this event had been experienced before. Thus, the automation aftereffects observed were minor compared to previous studies performed in driving simulators. The observed minor aftereffects seems to be independent of automation durations below 15 minutes. The extent to which the test environment alone (i.e. test track vs driving simulator) may explain this observation, requires further investigation. Independently of the test environment, however, the preparation-action-time consequence (i.e. the relation between the timings for the prompt; TOR; and conflict onset) is an important factor for explaining automation aftereffects. When the TOR is triggered before the conflict onset (Figure 2, bottom row) drivers are given time to put hands on wheel, feet on the pedals, and deactivate automation before being presented with the conflict. Consequently, these drivers may then able to respond to a safety-relevant event without a delay and perform similarly as when the same event is encountered in manual driving. Thus, unsupervised automation is not always associated with significant aftereffects compared to manual driving. Further, a 14-minute automation duration may result in that drivers require multiple attempts to deactivate automation, in contrast to a 4.5-minute duration where one attempt may be sufficient. It is therefore important to design a human-machine interface that facilitates an easy and intuitive procedure for drivers to deactivate automation, since some drivers may be particularly challenged after a longer automation duration.

Implications on safe vehicle automation

This thesis demonstrates that automation does not always result in detrimental automation effects in terms of unsafe driver response and performance in safety-relevant events in realistic settings (on test track with real vehicles). In fact, most drivers (72% of the drivers in Paper I and 100% of the drivers in Paper II) were able to perform safe manual intervention and driving performance after a period of assisted driving and unsupervised automation. However, this thesis also demonstrates that, after having supervised an assisted driving system for thirty minutes, some drivers (28% of drivers in Paper I) may not understand the need to act in a safety-relevant event if it is not preceded by a system-prompt. These drivers may respond late and crash, despite having eyes on the threat and being aware of the imminent crash. Thus, assisted driving without prompts is a topic that requires further research to understand how to make sure that drivers understand their role and responsibility to act when system limitations occur. For example, more work is needed to understand if a prompt prior to the safety-relevant event would result in appropriate driver response. However, since prompts may not always be an alternative when system limitations occur, there is also a need to investigate the benefits of different supervision strategies (e.g. hands-on-wheel requirements that require active steering input). Further, the safety of vehicle automation can be objectively measured through investigations of the driver response process as safety-relevant events occur. Through such investigations, this thesis demonstrates that safe vehicle automation may depend on the ability of automation to prompt drivers before the safety-relevant event. However, as highlighted by Paper II the timing of this prompt (in relation to the conflict onset) is also important. Finally, more work is needed to understand how automation aftereffects may be a consequence of the preparation-action-time (i.e. the time drivers need to become ready-to-act after automation) or by an underlying human cognitive mechanism (e.g. reduced situation awareness).

References

Bainbridge, L. (1983). Ironies of automation. *Automatica, 19*(6), 775-779. https://doi.org/10.1016/0005-1098(83)90046-8

Bianchi Piccinini, G., Lehtonen, E., Forcolin, F., Engström, J., Albers, D., Markkula, G., ... & Sandin, J. (2019). How do drivers respond to silent automation failures? Driving simulator study and comparison of computational driver braking models. *Human factors*, 0018720819875347.

Bolstad, C. A., Cuevas, H. M., Wang-Costello, J., Endsley, M. R., & Angell, L. S. (2010). Measurement of situation awareness for automobile technologies of the future. *Performance metrics for assessing driver distraction: the quest for improved road safety*, 195-213.

Bourrelly, A., de Naurois, C. J., Zran, A., Rampillon, F., Vercher, J. L., & Bourdin, C. (2019). Long automated driving phase affects take-over performance. *IET Intelligent Transport Systems*, *13*(8), 1249-1255.

Bärgman, J., Boda, C. N., & Dozza, M. (2017). Counterfactual simulations applied to SHRP2 crashes: The effect of driver behavior models on safety benefit estimations of intelligent safety systems. *Accident Analysis & Prevention*, *102*, 165-180.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, *36*(3), 181-204.

De Winter, J. C., Happee, R., Martens, M. H., & Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation research part F: traffic psychology and behaviour*, *27*, 196-217.

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*, *37*(1), 32–64. https://doi.org/10.1518/001872095779049543

Endsley, M. R. (2015). Situation Awareness Misconceptions and Misunderstandings. *Journal of Cognitive Engineering and Decision Making*, *9*(1), 4–32. https://doi.org/10.1177/1555343415572631

Endsley, M. R., & Kiris, E. O. (1995). The Out-of-the-Loop Performance Problem and Level of Control in Automation. *Human Factors*, *37*(2), 381–394. https://doi.org/10.1518/001872095779064555

Engström, J., Bärgman, J., Nilsson, D., Seppelt, B., Markkula, G., Piccinini, G. B., & Victor, T. (2018). Great expectations: a predictive processing account of automobile driving. *Theoretical issues in ergonomics science*, *19*(2), 156-194.

Eriksson, A., & Stanton, N. A. (2017). Takeover Time in Highly Automated Vehicles: Noncritical Transitions to and From Manual Control. *Human Factors*, *59*(4), 689–705. https://doi.org/10.1177/0018720816685832

Euro NCAP. (2018) *2018 Automated Driving Tests.* Retrieved from https://www.euroncap.com/en/vehicle-safety/safety-campaigns/2018-automated-driving-tests/

Feldhütter, A., Gold, C., Schneider, S., & Bengler, K. (2017). How the duration of automated driving influences take-over performance and gaze behavior. *Advances in ergonomic design of systems, products and processes* (pp. 309-318). Springer, Berlin, Heidelberg.

Fisher, D. L., Rizzo, M., Caird, J., & Lee, J. D. (2011). *Handbook of driving simulation for engineering, medicine, and psychology*. CRC Press.

Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013). "Take over!" How long does it take to get the driver back into the loop?. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 57, No. 1, pp. 1938-1942).

Gustavsson, P., Victor, T. W., Johansson, J., Tivesten, E., Johansson, R., & Ljung Aust, M. (2018). What were they thinking? Subjective experiences associated with automation expectation mismatch. *Proceedings of the 6th Driver Distraction and Inattention conference, Gothenburg, Sweden* (pp. 15-17).

Habibovic, A., Andersson, J., Nilsson, M., Lundgren, V. M., & Nilsson, J. (2016). Evaluating interactions with non-existing automated vehicles: three Wizard of Oz approaches. *IEEE Intelligent Vehicles Symposium (IV)* (pp. 32-37). IEEE.

Happee, R., Gold, C., Radlmayr, J., Hergeth, S., & Bengler, K. (2017). Take-over performance in evasive manoeuvres. *Accident Analysis & Prevention*, *106*, 211-222.

ISO. (2018). *Road Vehicles: Human Performance and State in the Context of Automated Driving: Part 1 – Terms and Definitions* (Standard No. 21959). Retrieved from: https://www.iso.org/standard/73927.html

Jarosch, O., & Bengler, K. (2018). Is It the Duration of the Ride or the Non-driving Related Task? What Affects Take-Over Performance in Conditional Automated Driving?. *Congress of the International Ergonomics Association* (pp. 512-523). Springer, Cham.

Kelley, J. F. J. (2018). Wizard of Oz (WoZ): a yellow brick journey. *Journal of Usability Studies*, *13*(3), 119-124.

Kruschke, J.K., Liddell, T.M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin and Review*.

Körber, M., Baseler, E., & Bengler, K. (2018). Introduction matters: Manipulating trust in automation and reliance in automated driving. *Applied ergonomics*, *66*, 18-31.

Larsson, A. F., Kircher, K., & Hultgren, J. A. (2014). Learning from experience: Familiarity with ACC and responding to a cut-in situation in automated driving. *Transportation research part F: traffic psychology and behaviour*, *27*, 229-237.

Lee, J. D., McGehee, D. V., Brown, T. L., & Reyes, M. L. (2002). Collision Warning Timing, Driver Distraction, and Driver Response to Imminent Rear-End Collisions in a High-Fidelity Driving Simulator. *Human Factors*, *44*(2), 314–334. https://doi.org/10.1518/0018720024497844

Lee, J. D., Wickens, C. D., Liu, Y., & Boyle, L. N. (2017). *Designing for people: An introduction to human factors engineering*. CreateSpace.

Llaneras, R. E., Cannon, B. R., & Green, C. A. (2017). Strategies to assist drivers in remaining attentive while under partially automated driving: Verification of human–machine interface concepts. *Transportation research record*, *2663*(1), 20-26.

Louw, T., Merat, N., & Jamson, A. H. (2015). Engaging with highly automated driving. To be or not to be in the loop. *8th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (pp. 190–196).

Macadam, C. C. (2003). Understanding and modeling the human driver. *Vehicle system dynamics*, *40*(1-3), 101-134.

McDonald, A. D., Alambeigi, H., Engström, J., Markkula, G., Vogelpohl, T., Dunne, J., & Yuma, N. (2019). Toward Computational Simulations of Behavior During Automated Driving Takeovers: A Review of the Empirical and Modeling Literatures. *Human Factors*, *61*(4), 642–688. https://doi.org/10.1177/0018720819829572

McLaughlin S., Hankey J., Dingus T. (2009). *Driver Measurement: Methods and Applications.* Engineering Psychology and Cognitive Ergonomics. Springer.

Merat, N., Seppelt, B., Louw, T., Engström, J., Lee, J. D., Johansson, E., ... & McGehee, D. (2019). The "out-of-the-loop" concept in automated driving: Proposed definition, measures and implications. *Cognition, Technology & Work*, *21*(1), 87-98.

Michon, J.A. (1985). A Critical Review of Driver Behaviour Models: What Do We Know? What Should We Do?. *Human Behaviour and Traffic Safety*.

Mole, C. D., Lappi, O., Giles, O., Markkula, G., Mars, F., & Wilkie, R. M. (2019). Getting back into the loop: the perceptual-motor determinants of successful transitions out of automated driving. *Human factors*, 0018720819829594.

Morando, A. (2019). *Drivers' Response to Attentional Demand in Automated Driving*. Division of Vehicle Safety, Department of Mechanics and Maritime Sciences, Chalmers University of Technology.

Morando, A., Victor, T., Bengler, K., & Dozza, M. (2020). Users' Response to Critical Situations in Automated Driving: Rear-Ends, Sideswipes, and False Warnings. *IEEE Transactions on Intelligent Transportation Systems*.

National Highway Traffic Safety Administration. (2015). *Critical reasons for crashes investigated in the National Motor Vehicle Crash Causation Survey*. Retrieved from: https://crashstats.nhtsa.dot.gov /Api/Public/Publication/812115

National Transportation Safety Board. (2017). *Collision Between a Car Operating With Automated Vehicle Control Systems and a Tractor-Semitrailer Truck Near Williston, Florida, May 7, 2016*. Highway Accident Report. Retrieved from: https://www.ntsb.gov/investigations/accidentreports/pages/har1702.aspx

National Transportation Safety Board. (2018). *PRELIMINARY HIGHWAY REPORT HWY18FH011.* Retrieved from: https://www.ntsb.gov/investigations/accidentreports/ pages/hwy18fh011-preliminary.aspx

National Transportation Safety Board. (2019). *PRELIMINARY HIGHWAY REPORT HWY19FH008.* Retrieved from: https://www.ntsb.gov/investigations/accidentreports/ pages /hwy19fh008-preliminary-report.aspx

Naujoks, F., Purucker, C., Neukum, A., Wolter, S., & Steiger, R. (2015). Controllability of partially automated driving functions–does it matter whether drivers are allowed to take their hands off the steering wheel?. *Transportation research part F: traffic psychology and behaviour*, *35*, 185-198.

Naujoks, F., Purucker, C., Wiedemann, K., Neukum, A., Wolter, S., & Steiger, R. (2017). Driving performance at lateral system limits during partially automated driving. *Accident Analysis & Prevention*, *108*, 147-162.

Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2014). Human performance consequences of stages and levels of automation: an integrated meta-analysis. *Human factors*, *56*(3), 476–488. https://doi.org/10.1177/0018720813501549

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors*, *39*(2), 230–253.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, *30*(3), 286-297.

Payre, W., Cestac, J., & Delhomme, P. (2016). Fully automated driving: Impact of trust and practice on manual control recovery. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *58*(2), 229–241.

SAE. (2018). SAE J3016 taxonomy and definitions for terms related to driving automation systems for on-road motor vehicle. Retrieved from: https://www.sae.org/standards/content/j3016_201806/

Seppelt, B. D., & Victor, T. W. (2016). Potential solutions to human factors challenges in road vehicle automation. In *Road vehicle automation 3* (pp. 131-148). Cham, Switzerland: Springer International.

Strand, N., Nilsson, J., Karlsson, I., & Nilsson, L.B. (2014). Semi-automated versus highly automated driving in critical situations caused by automation failures. *Transportation Research Part F-traffic Psychology and Behaviour*, *27*, 218-228.

Thatcham Research. (2018). *Assisted and Automated Driving: Definition and Assessement: Summary Document.* Retrieved from: https://news.thatcham.org/documents/assisted-and-automated-definition-and-assessment-summary-79399

Thatcham Research. (2019). *Defining safe automated driving: Insurer requirements for highway automation*. Retrieved from: https://www.thatcham.org/what-we-do/automated-driving

United Nations Economic Commission for Europe. (2017). *Addendum 78: UN Regulation 79.* Retrieved from: https://www.unece.org/fileadmin/DAM/trans/main/wp29/.../R079r4e.pdf

United Nations Economic Commission for Europe. (2020). *Proposal for a new UN Reguation on: uniform provisions concerning the approval of vehicles with regard to Automated Lane Keeping Systems.* Retrieved from: https://www.unece.org/trans/main/wp29/wp29wgs/wp29grva/grvainf20200.html

University of Michigan. (2018). *Sampling from Well Defined Populations* [Online Video]. Retrieved from: https://www.coursera.org/learn/understanding-visualization-data/lecture/uggvz/sampling-from-well-defined-populations

Victor, T., Dozza, M., Bärgman, J., Boda, C. N., Engström, J., Flannagan, C., ... & Markkula, G. (2015). *Analysis of naturalistic driving study data: Safer glances, driver inattention, and crash risk* (No. SHRP 2 Report S2-S08A-RW-1).

Victor, T. W., Tivesten, E., Gustavsson, P., Johansson, J., Sangberg, F., & Ljung Aust, M. (2018). Automation expectation mismatch: Incorrect prediction despite eyes on threat and hands on wheel. *Human factors*, *60*(8), 1095-1116.

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human factors*, *50*(3), 433–441.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, *3*(2), 159-177.

Wickens, C. D., Hooey, B. L., Gore, B. F., Sebok, A., & Koenicke, C. S. (2009). Identifying black swans in NextGen: Predicting human performance in off-nominal conditions. *Human Factors*, *51*(5), 638-651.

World Health Organization. (2018). *Global status report on road safety 2018*. World Health Organization.

Young, M. S., & Stanton, N. A. (2002). Malleable attentional resources theory: a new explanation for the effects of mental underload on performance. *Human factors*, *44*(3), 365-375.

Young, M. S., & Stanton, N. A. (2007). Back to the future: brake reaction times for manual and automated vehicles. *Ergonomics*, *50*(1), 46–58.

Zeeb, K., Buchner, A., & Schrauf, M. (2016). Is take-over time all that matters? The impact of visual-cognitive load on driver take-over quality after conditionally automated driving. *Accident; analysis and prevention*, *92*, 230–239.

Zhang, B., de Winter, J., Varotto, S., Happee, R., & Martens, M. (2019). Determinants of take-over time from automated driving: A meta-analysis of 129 studies. *Transportation research part F: traffic psychology and behaviour*, *64*, 285-307.