CHEMNITZ UNIVERSITY
OF TECHNOLOGY

Drivers' reliance on lane keeping assistance systems as a function of the level of assistance

Dissertation

zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)

vorgelegt der Fakultät für Human- und Sozialwissenschaften der
Technischen Universität Chemnitz

von Anke Popken (geb. Mogilka), geboren am 27.04.1980 in Karl-Marx-Stadt

Betreuer: Prof. Dr. Josef F. Krems

Chemnitz, den 13.05.2009

URL: http://archiv.tu-chemnitz.de/pub/2010/0048

Drivers’ reliance on lane keeping assistance systems as a function of the level of assistance

ACKNOWLEDGEMENTS

This driving simulator study was conducted in the framework of the European Network of Excellence HUMANIST (HUMAN centred design for Information Society Technologies) and was supported by a 3-years PhD grant and by a grant for infrastructure-sharing between network partners from the European Commission.

I would like to thank my supervisor, Prof. Dr. Josef F. Krems for his continuous support over the time of my PhD. My special thanks go to Dr. Lena Nilsson for her great engagement in this study, her scientific support, and for giving me such a warm reception at VTI. It was a joy to work with you!

Great thanks to the following people at VTI (in alphabetical order) who invested a lot of time and effort in rendering it possible to carry out this study and who made me have such a great time in Sweden: Mikael Adlers, Anne Bolling, Håkan Jansson, Laban Källgren, Albert Kircher, Katja Kircher, Göran Palmkvist, Beatrice Söderström, Jerker Sundström, Birgitta Thorslund, and Jan Törnros. Special thanks to Björn Peters for his scientific support during the last stage of my PhD.

Last but not least, my great thanks to the following student assistants for their support with the seemingly infinite ☺ data analysis: Nina Bär, Juliane Haupt, and Claudia Ihle.

This work wouldn’t have been possible without the constant support by my husband, Markus Popken. Thank you for all!

This work is dedicated to my grandfather.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ACC	Adaptive Cruise Control
ADAS	Advanced Driver Assistance Systems
HC	Heading Control
ICT	Information and Communication Technology
ITS	Intelligent Transport Systems
LDW	Lane Departure Warning
NA	No Assistance
NASA-TLX	NASA Task Load Index
RSME	Rating Scale Mental Effort
RTLX	Raw Task Load Index
SA	Situation Awareness
SDT	Signal Detection Theory
SSS	Stanford Sleepiness Scale
THW	Time Headway
TLC	Time to Line Crossing
TTC	Time to Collision

SUMMARY

Advanced Driver Assistance Systems (ADAS) are increasingly introduced into today's modern vehicles with the aim to support the driver while driving, to reduce driver errors and thereby to increase traffic safety. However, concerns are raised that drivers may adapt to the increasing automation of the driving task in unanticipated ways. A major concern in this regard is a reduced involvement of drivers in the driving task as response to the changing task demands. Since a complete reliable functioning of these systems cannot be guaranteed in any condition, the driver must be prepared to intervene appropriately in case of a system malfunction or failure.

One major aim of the current study was to investigate changes in drivers' active engagement in the driving task as a function of the increasing automation of the lateral control task. Changes in drivers' task engagement were studied by referring to two established theoretical concepts: drivers' reliance (on a lane keeping assistance system) and drivers' Situation Awareness (SA). Reliance on automation was defined as the deliberate act of operators to allocate varying levels of control to an automated system. High reliance on automation is associated with operators' less active engagement in the automation-controlled task and with a neglect of critically monitoring and verifying the appropriateness of the automated system's actions. In this study, drivers' reliance on a lane keeping assistance system was measured by their preparedness to divert their visual attention away from the road scene during their performance of a visually demanding secondary task.

The study was carried out in the VTI advanced moving-base driving simulator III in Linköping, Sweden. Forty-five Swedish drivers (25 male, 20 female) aged between 25 and 46 years (*M* = 33.7; *SD* = 6.4) drove on a simulated rural road of 65 km length with one of three levels of lane keeping assistance: (1) with a high level of lane keeping assistance realised by a Heading Control (HC) system, (2) with a low level of lane keeping assistance realised by a Lane Departure Warning (LDW) system, and (3) without lane keeping assistance (No Assistance, NA). During both experimental sessions drivers had to perform a visually demanding secondary task while driving. In Session 2, drivers additionally encountered eight critical driving situations. Seven of these situations were "safety-critical" and required drivers to react in order to avoid potential collisions with other traffic. The timeliness and abruptness of drivers' reaction to these situations served as performance-based measures of their SA. The critical situations always occurred in periods of concurrent secondary task performance.

It was hypothesised that drivers' reliance on a lane keeping assistance system increases the more completely the system automates the lateral control task. The results

did only partially support this hypothesis. Only one subgroup of HC drivers was found to rely to a large extent on the HC system, whereas another subgroup of HC drivers refused to rely on the HC system at all. Drivers in both subgroups differed significantly in the mean duration of their single glances away from the road scene during performance of the secondary task. Drivers who relied to a large extent on the HC system glanced considerably longer away from the road scene than LDW drivers and NA drivers; whereas drivers who did not rely on the HC system at all had considerably shorter eyes-off-the-road times than LDW and NA drivers. LDW drivers did not differ from NA drivers in their visual attention allocation strategies, indicating that they did not rely on the LDW system.

An analysis of factors that were assumed to influence drivers' reliance on the lane keeping assistance systems revealed that drivers' trust in the systems and their energetic arousal were able to predict a significant proportion of the variance in drivers' reliance on the HC system. However, those two factors had no influence on drivers' reliance on the LDW system. The greater drivers' trust in the HC system and the lower their activation level, the more did they rely on the HC system. Drivers' trust and arousal together accounted for about 42% of the variance in drivers' reliance on the HC system. Drivers' evaluation of the HC system's competence (in preventing them from departures from the driving lane) was found to be the strongest predictor of their trust in the HC system. In addition, individual driver characteristics (driving style) had a significant influence on drivers' trust in the HC and in the LDW system.

A second major aim of the study was to investigate changes in drivers' reliance in response to changes in task demands that affected the performance of the lane keeping assistance systems. According to the predictions, drivers' reliance on the lane keeping assistance systems significantly decreased in response to their experience of critical driving situations that partly represented functional limits of the lane keeping assistance systems. Although drivers' experience of functional system limits affected their perception of the lane keeping assistance systems, drivers' trust in the systems did not decline. In response to the changing driving task demands drivers' effort investment in the driving task and their workload significantly increased. There was however no indication that those effects were more pronounced the greater the lane keeping assistance systems' interventions in driving.

No consistent effect of the level of lane keeping assistance on drivers' reactions to the critical driving situations was found. It is suggested that changes in drivers' cognitive processes may not become apparent before a longer time of drivers' interaction with the systems. Results further indicated that especially drivers' interaction with the HC system required substantial conscious control in order to deal with the system's continuous interventions in driving. It is suggested that the results of this study should be complemented by long-term studies in order to observe the relative impact of different types of human adaptation processes on drivers' active engagement in the driving task over prolonged periods of drivers' interaction with driver assistance systems.

ZUSAMMENFASSUNG

Fahrerassistenzsysteme (Advanced Driver Assistance Systems, ADAS) werden zunehmend in heutige Fahrzeuge eingebaut mit dem Ziel, den Fahrer beim Fahren zu unterstützen, Fahrfehler zu reduzieren und damit die Fahrsicherheit zu erhöhen. Es bestehen jedoch Bedenken, dass Fahrer in unerwarteter Weise an die zunehmende Automatisierung der Fahraufgabe adaptieren. In besonderem Maße wird befürchtet, dass der Fahrer sich durch die verändernden Aufgabenanforderungen aus der Fahraufgabe zurückzieht und es dadurch zu einer nachlassenden Involviertheit des Fahrers in die Fahraufgabe kommt. Da eine vollständige Funktionszuverlässigkeit von Assistenzsystemen nicht unter allen Umständen garantiert werden kann, muss gewährleistet sein dass der Fahrer in angemessener Weise auf potentielle Systemfehler oder Systemausfälle reagieren kann.

Ein Hauptziel der vorliegenden Arbeit bestand darin zu untersuchen, inwieweit sich die Involviertheit des Fahrers in die Fahraufgabe mit zunehmender Automatisierung der Querführung verändert. Um dies zu untersuchen wurde auf zwei theoretische Konzepte Bezug genommen: auf das Verlassen (reliance) der Fahrer auf die entsprechenden Spurhalteassistenzsysteme und auf das Situationsbewusstsein (Situation Awareness) der Fahrer. Verlassen auf Automatisierung wurde definiert als die willentliche Übertragung eines veränderlichen Ausmaßes an Kontrolle an ein System. Ein starkes Verlassen auf Automatisierung steht in Zusammenhang mit dem Rückzug des Menschen aus der teilweise automatisierten Aufgabe (oder dem Prozess), verbunden mit der mangelhaften Überwachung und Überprüfung von Systemhandlungen. In dieser Studie wurde das Verlassen der Fahrer auf ein Spurhalteassistenzsystem operationalisiert durch die Bereitschaft der Fahrer, während der Bearbeitung einer visuell beanspruchenden Zweitaufgabe ihre visuelle Aufmerksamkeit von der Straße abzuwenden.

Die vorliegende Studie wurde im Fahrsimulator mit Bewegungsplattform bei VTI in Linköping, Schweden, durchgeführt. 45 schwedische Fahrer im Alter zwischen 25 und 46 Jahren (*M* = 33.7, *SD* = 6.4) fuhren auf einer simulierten Landstraße von 65 km Länge mit je einem von drei Assistenzgraden: (1) mit einem hohen Grad an Spurhalteassistenz realisiert durch ein Heading Control (HC) System, (2) mit einem mittleren Grad an Spurhalteassistenz realisiert durch ein Lane Departure Warning (LDW) System, oder (3) ohne Spurhalteassistenz (No Assistance, NA). Während beider experimentellen Sessions hatten die Fahrer eine visuell beanspruchende Zweitaufgabe während der Fahrt zu bearbeiten. In Session 2 wurden die Fahrer zusätzlich mit acht kritischen Fahrsituationen konfrontiert. Sieben dieser Fahrsituationen waren sicherheitskritisch und erforderten eine Reaktion seitens der Fahrer um potentielle Kollisionen mit anderen Verkehrsteilnehmern zu vermeiden. Die Rechtzeitigkeit und die Abruptheit der Reaktionen der Fahrer auf diese Situationen wurden als performanzbasierte Maße von Situation Awareness verwendet.

Es wurde erwartet, dass das Verlassen der Fahrer auf ein Spurhalteassistenzsystem mit steigendem Automatisierungsgrad der Fahraufgabe zunimmt. Die Ergebnisse bestätigten diese Hypothese nur teilweise. Vielmehr zeigte sich, dass nur eine Subgruppe der HC Fahrer sich in starkem Maße auf das HC System verließ, wohingegen ein anderer Teil der HC Fahrer sich überhaupt nicht auf das System verließ. Die Fahrer in beiden Subgruppen unterschieden sich signifikant in ihrer mittleren Blickabwendungsdauer von der Straße während der Bearbeitung der Zweitaufgabe. Fahrer die sich in starkem Maße auf das HC System verließen, wandten ihre visuelle Aufmerksamkeit bedeutend länger von der Straße ab als LDW Fahrer und NA Fahrer; wohingegen Fahrer, die sich nicht auf das HC System verließen, erheblich kürzere Blickabwendungen von der Straße riskierten als LDW Fahrer und NA Fahrer. LDW Fahrer unterschieden sich in ihren Blickstrategien nicht von NA Fahrern (Kontrollgruppe), was darauf schließen lässt dass sie sich nicht auf das LDW System verließen.

Eine Analyse der angenommenen Einflussfaktoren auf das Verlassen der Fahrer auf die Spurhalteassistenzsysteme ergab, dass das Vertrauen der Fahrer in die Systeme und ihr Aktivierungsniveau (arousal) signifikant zur Vorhersage des Ausmaßes des Verlassens der Fahrer auf das HC System beitrugen, nicht aber zur Vorhersage des Ausmaßes des Verlassens der Fahrer auf das LDW System. Je größer das Vertrauen der Fahrer in das HC System und je geringer ihr Aktivierungsniveau, desto stärker verließen sich die Fahrer auf das HC System. Diese beiden Faktoren erklärten zusammen 42% der Varianz im Verlassen der Fahrer auf das HC System. Die Einschätzung der Systemkompetenz (bezüglich dessen Fähigkeit ein Abkommen der Fahrer von der Fahrspur zu vermeiden) zeigte sich als stärkster Prädiktor für das Vertrauen der Fahrer in das HC System. Darüber hinaus hatten individuelle Eigenschaften der Fahrer (Fahrstil) einen signifikanten Einfluss auf das Vertrauen der Fahrer in die Spurhalteassistenzsysteme.

Ein zweites Ziel der Studie bestand darin zu untersuchen, inwieweit verändernde Situationsanforderungen (die die Performanz der Systeme beeinträchtigten) sich auf das Verlassen der Fahrer auf die Assistenzsysteme auswirkten. Entsprechend den Vorhersagen nahm das Verlassen der Fahrer auf die Assistenzsysteme signifikant ab in Folge ihres Erlebens kritischer Fahrsituationen, welche teilweise funktionale Grenzen der Spurhalteassistenzsysteme widerspiegelten. Obwohl das Erleben kritischer Fahrsituationen sich auf die Systembewertung auswirkte, nahm das Vertrauen der Fahrer in die Systeme nicht ab. In Folge der wachsenden Aufgabenanforderungen investierten die Fahrer signifikant mehr Ressourcen in die Fahraufgabe, und ihre subjektive Beanspruchung nahm signifikant zu. Entgegen der Erwartungen gab es jedoch keine Hinweise darauf, dass diese Effekte ausgeprägter waren je stärker das System in die Spurhaltung eingriff.

Es wurden keine konsistenten Belege für einen Effekt des Assistenzgrades auf die Reaktionen der Fahrer in kritischen Fahrsituationen gefunden, welche auf eine Beeinträchtigung des Situationsbewusstseins der Fahrer hinweisen würden. Es wird vermutet, dass mögliche Veränderungen in kognitiven Prozessen sich nicht vor einer länger andauernden Interaktion der Fahrer mit Fahrerassistenzsystemen zeigen. Die Ergebnisse deuten weiterhin darauf hin, dass vor allem die Interaktion der Fahrer mit dem HC System ein erhebliches Ausmaß an bewusster Kontrolle und an Lernaufwand erforderte um mit den kontinuierlichen Lenkeingriffen des Systems zurechtzukommen. Es wird vorgeschlagen, die Ergebnisse dieser Studie durch Langzeituntersuchungen zu ergänzen um den relativen Einfluss verschiedener Adaptationsprozesse in Folge der zunehmenden Automatisierung der Fahraufgabe auf die Involviertheit des Fahrers in die Fahraufgabe über längere Zeit der Interaktion des Fahrers mit Fahrerassistenzsystemen zu untersuchen.

1 INTRODUCTION

The increase in mobility and in the use of road transport poses serious safety and environmental challenges on the society at the beginning of the 21st century. In Europe, the number of cars per thousand persons has increased from 232 in 1975 to 464 in 2007. Besides economic and environmental problems (congestion, fuel consumption, emissions), the high number of accidents caused traffic safety to be a key topic on the agenda of the European Commission's actions. In 2007, the number of accidents involving personal injuries was equal to 1,276,800 in Europe, involving 42,448 fatalities (European Commission, 2009). In 2001, the European Commission issued a White Paper with the ultimate objective to halve the number of road traffic injuries by 2010 (European Commission, 2001). As partial fulfilment of this objective, in 2005, the Commission has launched the Intelligent Car Initiative "Raising Awareness of Information and Communication Technologies (ICT) for smarter, safer and cleaner vehicles" (European Commission, 2006). Such Intelligent Transport Systems (ITS) include various categories of system applications making use of the latest developments in information and communication technologies to interact between the car driver, the vehicle and the road environment as well as between vehicles (vehicle-to-vehicle communication) and between vehicle and road infrastructure such as traffic management systems (Archer, 2000). Thus, ITS are envisaged to have the potential to increase the overall efficiency and the productivity of the transport system, to use energy efficiently, to reduce environmental pollution and to improve traffic safety.

Improving traffic and driver safety is the ultimate objective of a specific category of intelligent systems implemented in the vehicle termed "Advanced Driver Assistance Systems" (ADAS). ADAS aim at assisting the driver while driving by providing real-time feedback and (safety) warning and/or by directly intervening in driving within different time-horizons in order to prevent accidents, e.g., by initiating braking in order to avoid a collision or by automating driving subtasks such as headway maintenance or lane keeping. Driven by the fact that human (driver) error contributes to 93% of all accidents, and that it is considered as the main underlying cause in three-quarters of these cases (e.g., German In-Depth Accident Study), the rationale behind ADAS is to reduce driver stress and errors, increase driver comfort and thereby avoid accidents (e.g., Brookhuis & De Waard, 2005; Nilsson, 2005; Stanton & Young, 1998). Today a number of systems and functions are available on the market (see e.g., Floudas, Amditis, Keinath, Bengler, & Engeln, 2004), whereas the trend goes towards an increasing automation of functions and assistance in more and more complex driving situations such as lane change manoeuvres and crossings of intersections (e.g., Carsten, 2005; Janssen, Wierda, & Van der Horst, 1995; Marchau & Van der Heijden, 2000; Walker, Stanton, & Young, 2001).

This trend towards increasing automation of the driving task shows similarities with technological developments in other transport areas like aviation and air traffic control in the 1980s and 90s (Hancock & Parasuraman, 1992). There it became apparent that automation does not simply supplant human activity, but changes the nature of tasks which often leads to unanticipated consequences for human performance and underlying cognitive processes (Parasuraman & Mouloua, 1996; Parasuraman, Sheridan, & Wickens, 2000). Leading researchers have suggested to learn from automation issues in other transport areas and to take the human-factors insights, proactively rather than retroactively, into account in the design of ADAS (Hancock & Parasuraman, 1992). The "human factor" is widely acknowledged nowadays by manufacturers and researchers who foster a "human-centred" or "driver-centred" design of ADAS (e.g., Cacciabue & Martinetto, 2004; Goodrich & Boer, 2003; Hancock, Parasuraman, & Byrne, 1996; Tango & Montanari, 2006), albeit some persistent uncertainties regarding the application of concepts and methods (Cacciabue, 2006).

One major concern with regard to automation is the so-called "out-of-the-loop performance" often documented by the difficulty of human operators to detect system malfunctions or automation failures when they do not actively (manually) control a process (Endsley & Kiris, 1995; Kessel & Wickens, 1982; Metzger & Parasuraman, 2005; Molloy & Parasuraman, 1996; Parasuraman, Molloy, Mouloua, & Hilburn, 1996; Parasuraman, Molloy, & Singh, 1993; Parasuraman, Mouloua, & Molloy, 1996; Skitka, Mosier, & Burdick, 2000; Wickens & Kessel, 1981). This difficulty is generally attributed to a lack of involvement of the human operator in the automation-controlled processes inherent in the shift from active to supervisory control (Sheridan, 2006). A number of cognitive and motivational factors accompanying humans' adaptation to the changing task demands are hypothesised to contribute to out-of-the-loop performance, among them vigilance problems associated with excessive system monitoring demands or due to underload and boredom, an inaccurate or incomplete mental representation of the current situation (situation awareness), and overreliance on automation and inappropriate trust in system capabilities. Those adaptation effects are insofar problematic as, despite the sophistication of current technologies, systems often do not meet the requirement of working completely reliable in any type of expected and unforeseen condition characterising highly dynamic environments such as road traffic (Parasuraman et al., 2000).

In consideration of the increasing automation of the driving task, overreliance on a driver assistance system's actions, accompanied by an incomplete mental representation of the current driving situation (situation awareness) represent major safety concerns based on the requirement that drivers must be prepared to intervene appropriately to take over manual control in case of a system failure (Brookhuis & De Waard, 2005; Carsten & Nilsson, 2001; De Waard, Van der Hulst, Hoedemaeker, & Brookhuis, 1999; Nilsson, 2005; Stanton & Young, 1998). Given the time-horizon available in suddenly evolving or unexpected situations, the driver's readiness to react accordingly in a timely manner is a prerequisite for safe driving.

In general, reliance on a system is operationally defined as the extent to which a system is used by operators (e.g., Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003; Dzindolet, Pierce, Beck, & Dawe, 2002; Lee & Moray, 1992, 1994; Parasuraman & Riley, 1997; Riley, 1996). Rather than a discrete decision of engaging and disengaging automation, reliance reflects a graded process of operators allocating control between themselves and a system (Lee & See, 2004). A number of empirical studies have demonstrated the strong influence of system characteristics on operator reliance. The reliability of a system's performance was found to affect human trust in the system (Lee & Moray, 1992; Muir & Moray, 1996), which in turn guided reliance (Muir & Moray, 1996). In general, people tend to rely on systems they trust, and tend to reject systems they distrust. Furthermore, an inverse relationship was found between reliance on a system and operators' monitoring of its behaviour (Muir & Moray, 1996). It is hypothesised that high reliance on a system negatively affects operators' situation awareness (Sarter & Woods, 1991; Scerbo, 1996) because operators invest less effort in updating an accurate and complete mental representation of the components of a situation controlled by the system.

The main assumption in the present study is that the level of automation, i.e. the level of control the driver has in accomplishing driving functions, has a similar effect as a system's reliability on drivers' reliance because they both reduce the necessity for driver interventions. In this driving simulator study drivers' reliance on a driver assistance systems and their situation awareness is investigated as a function of the degree to which the system automates the drivers' lane keeping task. Thereby, other influencing factors that are assumed to mediate this relationship were taken into account.

2 THEORETICAL BACKGROUND

2.1 Automation and Changes in Operators' Level of Task Engagement

Automation can be defined as "a system that accomplishes (partially or fully) a function that was previously, or conceivably could be, carried out (partially or fully) by a human operator" (Parasuraman et al., 2000, p. 287). Automation of functions is rarely realised in an all-or-none fashion, but can rather be characterised by a continuum of levels corresponding to varying degrees of computer vs. human operator autonomy (Sheridan & Verplanck, 1978).

According to Parasuraman et al. (2000), automation can be applied to different stages of human information processing and action regulation. According to their model, the following four types of functions can be automated to different degrees: (a) information acquisition, (b) information analysis, (c) decision and action selection, and (d) action implementation. Thus, systems can be distinguished according to the degree to which they automate each of these four functions. For example, an automobile night vision enhancement system may support the driver by presenting an additional infrared-image of the road scene in order to ease the detection of pedestrians at night, corresponding to a high level of information acquisition automation. Alternatively, the system may not provide an extra-image of the roads scene, but solely draw the drivers' attention to pedestrians once detected by an internal image processing algorithm, e.g. by highlighting them in the windscreen or by presenting an auditory warning. This would correspond to low automation of information acquisition and high automation of information analysis.

Parasuraman et al. (2000) propose to use their model of types and levels of automation as a framework for the decision to which degree particular functions should be automated in a given task context. While it is undisputable that automation is mostly beneficial for human performance when it works properly, problems arise from the fact that automated systems are rarely completely reliable. Full automation not only requires the technical ability to deal with all types of known errors, but also the ability to handle unforeseen faults and events. This requirement goes beyond the ability of most intelligent systems. There has been a growing research interest in how humans adapt to automation, given that they have to be able to intervene and to resume manual control in cases of automation failures or malfunctions. Of major concern in this regard is the operator's declining level of task engagement as a major hypothesised contributor for the out-of-the-loop performance problem. Parasuraman et al.'s taxonomy and those of others (e.g., Endsley & Kaber, 1999) is based on the justified assumption that the more automation intervenes in the human control processes, the more difficulties human operators will have to stay "in the loop", i.e. to maintain a high level of task involvement. Automation applied to higher-level cognitive processes (such as information analysis and decision-making) is expected to be particularly harmful for the operator's ability to stay in the loop and to detect and react efficiently to automation malfunctions (Parasuraman & Wickens, 2008). Two concepts related to the degree of operators' involvement in a task when assisted by an automated system will be the focus in this work. The first one is operators' reliance on automation, and the second one is operators' situation awareness.

Reliance on automation will be defined here as the deliberate act of allocation varying levels of control to an automated system. Inherent in this conceptualisation of reliance is the assumption that to the extent an operator decides to allocate control to a system, she or he will be less likely to actively engage in the respective task (assisted by automation), will increasingly hand over responsibility for maintaining overall performance to the system, and will invest less effort in critically monitoring and examining the appropriateness of the system's actions or behaviour. Reliance thus refers to the operator's efforts to stay actively involved in a task when this task is automated to a lower or greater extent.

There are concerns that operators may increasingly rely on automation when a greater portion of a task is automated, also referred to as overreliance or misuse of automation (Parasuraman & Riley, 1997; Sheridan & Parasuraman, 2005). Such concerns are supported by empirical studies in which the reliability of an automated system was manipulated. Those studies have shown that operators were more likely to allocate control to a system, invested less effort in system monitoring, and had more difficulties in detecting system malfunctions the higher the reliability of the automated system (e.g., Muir & Moray, 1996). Thus, the reduced necessity for manual control or intervention in highly reliable and highly automated systems can be assumed to foster operators' reliance on automation.

The second cognitive psychological concept related to the degree of operators' involvement in a task is situation awareness. *Situation awareness* refers to the operator's maintenance of a coherent and comprehensive understanding (mental representation) of the current situation (Endsley, 2000b). The availability of a current mental representation of the situation is regarded as prerequisite for the operator's ability to detect and to react to unforeseen critical events such as automation failures. Based on cognitive theoretical frameworks it is predicted that automation interrupts the cognitive processes necessary to achieve and maintain situation awareness. Thus, it is assumed that higher levels of automation will negatively affect operators' situation awareness (Endsley, 1996; Parasuraman et al., 2000).

The aim of this PhD work is to comprehensively study changes in the level of drivers' involvement in the driving task as a function of the increasing automation of the lateral control subtask. Changes in drivers' involvement in the driving task are assessed by referring to two related psychological concepts: drivers' reliance (on lane keeping assistance systems) and their situation awareness. The present work is guided by a theoretical framework on the impact of automation on operators' reliance and situation awareness that is presented below. The framework extends already existing theoretical accounts by viewing reliance and situation awareness in a broader context of operators' adaptation processes in response to automation. A review of empirical studies revealed that human adaptation responses occur at different levels referring to cognitive, motivational and energetic processes. The theoretical framework links changes in operators' level of task engagement to changes in human cognitive, motivational and energetic processes in response to automation.

2.2 A Theoretical Framework of Operators' Adaptation to Automation

Empirical studies have shown that automation affects human cognitive, motivational, and energetic processes. Most research on the effects of automation on human performance focused on operators' detection of critical system parameters as the major requirement of supervisory control (Moray, 1986; Sheridan, 2006). In those studies, operators' monitoring performance under manual control was usually compared with their monitoring performance when portions or the complete task became automated. A major finding that was replicated in almost all of these studies is the impaired detection performance for critical system states under automated conditions compared to manual control. In interpretation of this finding a number of explanatory factors were hypothesised: complacency, vigilance decrement, underload, loss of situation awareness, excessive trust in automation capabilities, and overreliance – to name the most frequent ones. A major shortcoming in many of these studies can be seen however in the lacking efforts to measure these hypothesised influencing factors explicitly. At the same time, the impairment in monitoring performance has been used as an equivalent measure for almost all of the influencing variables listed above.

In Figure 1 a conceptual framework is presented that differentiates between various types of human adaptation processes in response to automation that in turn influence operators' reliance on automation. In line with Lee and See's (2004) model of reliance on automation, *reliance* is defined as a dynamic process of allocating varying levels of control to an automated system. It is assumed that operators who increasingly rely on an automated system invest less effort in system monitoring, are less actively engaged in the task (assisted by automation) and increasingly hand over responsibility for overall task performance to the system.

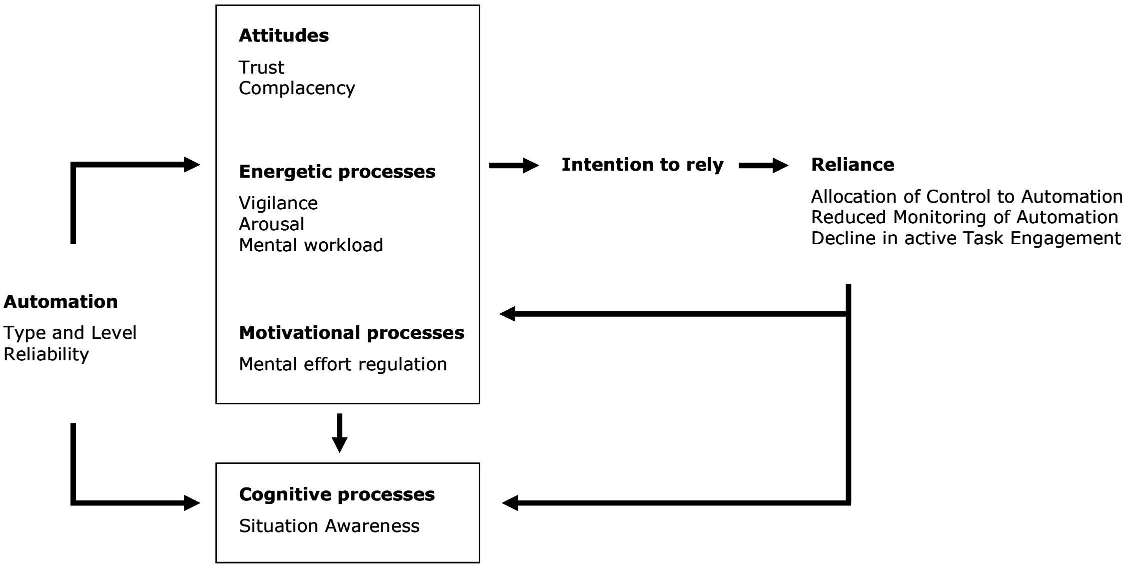


Figure 1. A conceptual framework of human adaptation processes in response to automation influencing operators' level of active task engagement, assessed by the two psychological concepts of reliance on automation and situation awareness.

Reliance is theoretically linked to two other concepts describing operators' adaptation to automation, namely complacency and automation bias. The state of these two concepts, but especially that of complacency, is still subject to an ongoing debate among researchers (see chapter 2.3.4.2). *Complacency* has been defined among other things as "self-satisfaction which may result in non-vigilance based on an unjustified assumption of satisfactory system state" (Parasuraman et al., 1993, p. 2, quoting NASA definition). The major problem with complacency lies in its nature as an inferred construct and the insufficient specification of observable behavioural indicators that would allow to draw conclusions about its existence (see Moray & Inagaki, 2000 for a critical review and suggestions for an operational definition). In fact, complacency appears to be a popular post-hoc explanation for a variety of observed human adaptation effects in response to automation without an attempt to verify its hypothesised causal effects. Although its explanational value seems to be restricted at present, complacency is included in the conceptual framework presented here. Complacency is treated as an attitude towards automation that influences reliance (as behaviour).

Automation bias refers to the operators' "tendency to use automated cues as a heuristic replacement for vigilant information seeking and processing" (Mosier & Skitka, 1996, p. 205). Rather than a unique concept related to humans' adaptation to automation, automation bias is here understood as an aspect or behavioural correlate of a high level of reliance related to the operators' reduced efforts to critically examine the actions of an automated system (cp. Bahner, Hüper, & Manzey, 2008).

Reliance on automation is viewed as a deliberate action, influenced by the operators' intention to rely (Lee & See, 2004). The intention to rely is mediated by humans' adaptation processes in response to automation. Research on automation revealed that automation affects operators' attitudes as well as human cognitive, motivational, and energetic processes. In line with theoretical assumptions and empirical evidence it is expected that operators' adaptive responses to automation increase in magnitude the more automation intervenes in the natural human control processes involved in task performance. Automation that relieves operators from higher cognitive processes is assumed to have a particular large impact on human cognitive, motivational and energetic processes (Parasuraman & Wickens, 2008).

As already mentioned, many studies have compared operators' performance under manual control conditions with their performance when particular aspects of a task became automated. Fewer studies have investigated the effects of varying levels or different types of automation on human adaptation processes and performance. Especially studies that focused on the effects of different levels of automation on operators' reliance are rare. Given the imperfections of today's technical systems in terms of reliability, researchers were especially interested in the effects of varying levels of automation reliability on reliance. This research however has offered important insights, and is used here to draw a parallel to the hypothesised effects of increasing levels of automation on operators' reliance. In particular, it is assumed that a high level of reliability and a high level of automation affect operators' reliance in a similar way, because they both reduce the necessity for manual control and for operators' active engagement in a task.

In line with other theoretical frameworks (Lee & See, 2004; Meyer, 2004), it is considered that adaptation processes are not solely determined by the type and level of automation and its performance, but also by situational demands and personality traits of the operator. *Situational demands* are expected to particularly affect the operators' energetic state (e.g., mental workload) and motivational processes. For example, the severity of consequences associated with potential automation failures is assumed to influence operators' subjective assessment of perceived risk, and in turn to guide their decision about mental effort investment (Fairclough, 2001) and their intention to rely on an automated system (Mosier & Skitka, 1996). A similar relationship is assumed for the influence of *personality traits* (such as sensation seeking and locus of control) on motivational processes (Brown & Noy, 2004; Jonah, 1997; Saad, 2004). Energetic processes are assumed to be influenced by individual differences in energetic arousal and resource availability (Matthews, Davies, & Lees, 1990), whereas individuals' propensity to trust technical systems is hypothesised to influence operators' attitudes towards automation and to cause interindividual variance in operators' trust that cannot be explained by differences in automation performance (Lee & See, 2004; Prinzel, Freeman, & Prinzel, 2005).

Based on cognitive theoretical frameworks it is assumed that automation also affects human information processing and the top-down and bottom-up control of attention (Shebilske, Goettl, & Garland, 2000). Higher levels of automation are therefore expected to interrupt the cognitive processes necessary to achieve and maintain a current and comprehensive mental representation of the situation (situation awareness). Because the continuous updating of the mental representation of the situation is regarded as a resource-intensive process, changes in operators' energetic state (such as decreasing vigilance) and motivational processes are expected to influence situation awareness (Endsley, 1996).

It can further be hypothesised that different types of adaptation processes in response to automation may influence reliance within different time horizons. Short-term changes in reliance may be particularly mediated by energetic and motivational processes susceptible to changes in task demands. Changes in cognitive processes are assumed to evolve not before a certain time period of continuous interaction with automation. Attitudes may influence reliance particularly at the beginning of operators' interaction with automation, and may later be replaced by learned action patterns and expectations regarding the automation's performance (cp. Riley, 1996).

The conceptual framework incorporates a feedback loop assuming that reliance on automation or the allocation of varying amounts of control to an automated system will subsequently affect energetic, motivational, and cognitive processes. Low reliance on automation coupled with the investment of substantial mental resources in system monitoring will over time likely affect the operator's energetic state (decline in vigilance) and motivational processes (switching to effort conservation strategies). Likewise, high reliance on automation may lead to a decrease in arousal, an increase in operators' trust and complacency (especially when reliance does not lead to negative performance consequences), and an increase in operators' tendency to adopt strategies of mental effort conservation. Reliance is also expected to have a negative influence on situation awareness due to the changes in cognitive processes associated with the increasing allocation of control to an automated system.

In the following chapters, theoretical foundations and empirical research underlying the conceptual framework will be reviewed in more detail. The following chapters are organised in two large parts. The first part deals with research on reliance on automation and the second part deals with research on situation awareness. Each part will begin by introducing the theoretical background of the respective concept. Afterwards, the focus is on empirical evidence regarding the impact of the automation level and/or automation reliability on both concepts. This review of empirical studies will be structured around the different types of human adaptation processes in response to automation in order to underline their mediating influence.

2.3 Reliance on Automation

2.3.1 What is reliance? Some critical preliminary remarks

Reliance has been defined above as the deliberate act of allocation varying levels of control to an automated system. Thus, reliance can be viewed as a continuum ranging from no reliance – specified as the neglect to allocate control to an automated system – to complete reliance – specified as the allocation of full control to an automated system.

In the research literature the term reliance is not always used consistently. Furthermore, there is no general accepted formal definition of reliance. Reliance on automation has been used as a synonym for *automation use*, or to refer to the degree of appropriateness of allocating varying levels control to an automated system. Beck, Dzindolet, and Pierce (2002) for example speak about "automation usage decisions" instead of reliance. In research on binary warning systems, the term reliance is often replaced by the term *dependence*, whereas two conceptually different types of operators' dependence on warning systems are termed *reliance* and *compliance* (e.g., Meyer, 2004; see chapter 2.3.3.2).

Parasuraman and Riley (1997) refer to the two opposite poles of the reliance continuum as automation disuse and misuse, incorporating a notion of inappropriateness of both a high and a low level of reliance with regard to the capabilities or the performance of the automation. Thus, misuse is specified as "too much" reliance, whereas disuse refers to "too little" reliance. According to Parasuraman and Riley, *misuse* can be understood as overreliance on automation and is usually conceptually linked to operators' excessive trust in system capabilities. *Overreliance* is described as uncritical reliance on an automated system (i.e. relying on a system when it does not work properly) associated with ineffective monitoring of its behaviour and impaired recognition of system limitations. *Disuse* on the opposite refers to the underutilisation or neglect of automated systems and their potential benefits by operators, often following operators' low acceptance or mistrust in automation capabilities, for instance due to a high false alarm rate. Whereas disuse of automation refers to not using automation although it would be beneficial for performance or efficiency, misuse refers to the "excessive" or inappropriate use of automation, i.e. using it "too much" or using it when it should not be used.

Unfortunately, in a number of studies one has contented oneself with interpreting operators' performance post-hoc as a sign of over-reliance, of inappropriate reliance, of operators having had excessive trust in automation capabilities, or of complacency. This yields particularly for studies in which operators were found to detect and react lately to system malfunctions under automated conditions compared to when they performed the same task manually. The post-hoc description of behaviours as a sign of inappropriate reliance or overreliance lacks however explanational value because it induces a circular reasoning: Why did an operator react lately to a system failure? Because she or he over-relied on the system. How does one know that she or he over-relied on the system? Because she or he reacted lately to a system failure. As pointed out in the last chapter, operators' impaired detection performance of critical system states under automated conditions has been attributed to a number of reasons with overreliance being only one of them. Furthermore, there is no a priori theoretical specification of "appropriate", and of when reliance switches from being appropriate to being inappropriate.

In conclusion, it appears important to distinguish between formal and operational definitions of reliance. From a theoretical point of view reliance on automation is not the same as automation use in the sense of binary decision of engaging vs. disengaging it (Gao & Lee, 2006; Lee & See, 2004). Operators' decisions to use automation or to manually perform a task may however be used as a measure for or as an operational definition of reliance. It is argued here however that reliance reflects a continuum along which operators dynamically allocate varying levels of control to an automated system. Operators may rely to different extents on an automated system when they use it or have engaged it. There are circumstances in which operators have no choice in using automation or not. Nevertheless, they may allocate different levels of control to automation which is expected to become apparent for example in the effort they put in critically monitoring the automation's actions. Thus, a high level of reliance on automation is theoretically linked to operators' less active engagement in the automation controlled processes and the handing over of responsibility for overall performance to the automation. It is noted however that the degree of reliance is related to the level of automation. A low level of reliance on high level automation (e.g., automation of action implementation) is associated with the allocation of a substantially larger amount of control to an automated system than a low level of reliance on low level automation (e.g., automation of information analysis). Thus, the range of the reliance continuum may shrink the higher the level of automation.

This conceptualisation of reliance demands for stronger process oriented measures of reliance which are able to depict strategic changes in operators' allocation of control to an automated system depending on the properties of the automation and situational demands. Indeed, models of operators' reliance on automation assume that reliance follows from a *dynamic* interaction among the operator, the automation, and the context.

Thus, the focus in the present study will be on drivers' strategic allocation of control to an assistance system. Drivers' reliance on the assistance system will be assessed via their preparedness to allocate their visual attention away from the road (the most relevant information channel while driving) as a function of the degree of assistance offered by the assistance system (level of automation) and drivers' experiences of functional system limits.

2.3.2 Theoretical approaches to reliance on automation

Theoretical models of operators' reliance on automated systems reveal that reliance depends on a *dynamic* interaction among the operator, system, and context (Dzindolet et al., 2002; Lee & See, 2004; Meyer, 2004; Riley, 1996). In all of these frameworks, the performance of a system forms the starting point for certain appraisal processes that constitute the basis for operators' decision to rely on the system. Operators' trust in the system has a major influence on their reliance and is based, more or less directly, on the performance of the system. Thus, all models assume that humans respond socially to automation by developing certain extents of trust in it; which in turn influence their reliance. Trust is proposed as an intervening variable that mediates reliance under conditions of uncertainty. The monitoring role of human operators in supervisory control implicates that operators have no complete knowledge of the exact properties of the automation in any possible conditions (Muir, 1994). Lee and See (2004) define trust as "the attitude that an agent will help achieve an individual's goal in a situation characterised by uncertainty and vulnerability" (p. 54). The researchers state further that "trust guides reliance when complexity and unanticipated situations make a complete understanding of the automation impractical" (Lee & See, 2004, p. 50). The models of reliance make partly different assumptions of how operators develop trust in automation and focus on different type of processes underlying reliance. They differ also in the level of detail used to specify those processes.

In an initial theory of operator reliance on automation, Riley (1996) assumes that the reliability of an automated system influences operators' trust in it. Reliance can be understood as the probability that an operator will use the system, and is influenced by operators' trust in it and their self-confidence in their own abilities to perform the task manually. Other factors, like task complexity, perceived workload and perceived risk, are hypothesised to influence operators' self-confidence and to intervene in the interaction of trust and self-confidence in determining reliance. It is assumed that if operators' trust in the automation exceeds their self-confidence in manual control, they most likely rely on automation. On the contrary, if operators are more confident in their own abilities than they trust the automation, then they will likely choose manual control. Empirical evidence for this was found in a number of studies (e.g., De Vries, Midden, & Bouwhuis, 2003; Lee & Moray, 1992, 1994; Lewandowsky, Mundy, & Tan, 2000). Riley's model also accounts - although not very detailed – for the influence of operator states (fatigue) on reliance and the time dynamics influencing the relation between trust and reliance. He assumes that with extended system use operators learn to recognise system states and to anticipate system behaviour; and that reliance is then more directly based on the outcomes of this learning process than on the mediating influence of trust. In consideration of the high individual differences in participants' reliance found in his experiments, Riley (1996) also assumes that peoples individual behaviour may be influenced only by a subset of the factors proposed in his model.

Dzindolet et al. (2002) elaborate to a greater extent the role of rational decision processes in determining reliance and the influence of biases on these processes (see also Beck et al., 2002). They also stress the important impact of social and motivational factors on reliance. Similar to Riley's (1996) model, Dzindolet et al. (2002) postulate a process in which the capabilities of the automation and the capabilities of the operator are weighed up against each other. However, contrary to Riley's (1996) model, this comparison is not influenced by operators' trust in the automation, but the outcome of the comparison itself forms the basis for operators' trust in the automation. In other words, trust follows the outcome of operators' rational judgement regarding the utility of the automation which is based on a comparison between the perceived reliability of the automation and the perceived reliability of manual operation. Certain biases act on the operators' perception of reliability (e.g., bias towards automation, self-serving biases). The notion of "perceived reliability" indeed is not trivial, as one can assume that the actual reliability of a system may never be known by the operator. Thus, the reliability of a system can only be inferred by subsequent operator interactions with the system. Dzindolet et al. (2002) predict that operators' reliance on automation is influenced by their trust in it and by other social and motivational factors, such as feelings of control and effort. In accordance with dynamics in group processes where social psychological phenomena like diffused responsibility and social loafing were found to play a role (Kerr & Bruun, 1983; Latane, Williams, & Harkins, 1979), they hypothesise that operators may feel less responsible for the consequences of performance when they collaborate on a task with an automated system than when they perform the task alone. According to Dzindolet et al., effort investment is based on the outcome of a cost-benefit analysis in which operator states such as fatigue and workload, operators' intrinsic interests in the task, as well as penalties and rewards associated with performance consequences, are taken into account. The difficulty of the task itself also has an impact on operators' effort in the way that operators are more prone to rely on a system if they perceive that task demands exceed their capabilities (e.g., Madhavan, Wiegmann, & Lacson, 2006).

Recently, Lee and See (2004) proposed a dynamic model of trust and reliance on automation. It offers an integrative framework for empirical research on trust in and reliance on automation and largely incorporates the hypothesised interactions among variables postulated in the two aforementioned models. Lee and See's (2004) model appears appealing for a number of reasons. First, it helps to clarify processes and interactions between variables because it distinguishes between beliefs, attitudes, intentions and behaviour. Thus, trust in automation is an attitude, and reliance is a behaviour. Trust is built on beliefs about the characteristics of the automation. Based on their level of trust, operators form the intention to rely on the automation. Once the intention is formed, other environmental (e.g., time constraints, task demands), individual (e.g., a person's predisposition to trust), organisational, and social factors may affect whether or not an operator actually relies on the automation. Second, Lee and See's model explains how people develop trust in automation and how system characteristics affect the appropriateness of trust. Trust is appropriate when it matches the (true) capabilities of the system. The design of the human-machine interface plays an important role in the calibration of trust. Third, the model considers certain dynamics in the relationship between trust and reliance. Trust is a nonlinear function of automation performance and the dynamic interaction between the operator and the automation. The performance of an automated system and its design, as well as situational demands and personality traits act together to affect operators cognitive and motivational processes intervening in the relationship between trust and reliance. Similar interactions between influencing factors are assumed by Meyer (2004) in his conceptual framework of operators' reliance on warning systems.

In summary, theoretical frameworks on reliance on automation suggest that operators are generally susceptible for the potential added value offered by automation. Operators are willing to allocate control to a system to the extent that this is perceived as being beneficial for performance and/or efficiency. The notion of effort appears to play a dominant role in operators' reliance on automation. Thus, more or less explicitly, theoretical frameworks assume that operators' degree of reliance is influenced by a utility judgement about the costs and benefits of allocating varying levels of control to automation. This utility judgement seems to be driven by two major factors: First, by the expected performance improvements associated with reliance; and second, by the potential to conserve mental effort through reliance on automation. High reliance is associated with less active task engagement and less intensive monitoring and verification of automation behaviour, and consequently, is considered as being less effortful and resource demanding than low reliance on automation. Furthermore, low reliance on automation can be associated with even considerably higher demands than manual control, because it generates the additional burden for operators to verify, correct or ignore the automation's actions. The performance and capabilities of the automation are assumed to have a major influence on operators' intention to rely on automation because they (in interaction with the environmental context and operator characteristics) affect energetic and motivational processes and operators' attitudes towards automation. Operators' trust in automation (as an attitude towards automation) is hypothesised to act as a decision heuristic in guiding reliance when operators do not have complete or accurate information about the automation's performance.

Up to now, empirical studies have almost exclusively focused on the impact of varying levels of automation reliability on reliance on automation. The major assumption in the current work is however that the level of automation has a similar effect as automation reliability on reliance because it is hypothesised to affect underlying energetic and motivational processes in a similar way.

In the following, empirical studies on the effects of automation reliability and the level of automation on reliance will be reviewed in more detail. Chapter 2.3.3 focuses on empirical evidence for the assumed relationship between automation performance (reliability) and operators' reliance on automation and examines potential manifestations of operators' overreliance on automation. The subsequent chapters 2.3.4 to 2.3.6 aim at demonstrating the mediating influence of operators' attitudes as well as of energetic and motivational processes on reliance.

2.3.3 The influence of automation reliability on operators' reliance on automation

A great deal of research on operators' reliance on automated systems has naturally focused on the impact of system properties - such as its performance - on reliance. Many of these studies have focused on operators' ability to detect automation failures or malfunctions as one major requirements of supervisory control. As completely reliable automation is rare, the demand for operators to rely 'appropriately' on automation has emerged, i.e. rely on automation when it is highly reliable, and take over manual control when it is not. Although this pattern of automation use can be considered fully rational in order to maximise overall system efficiency and safety, it seems not to be the way human operators generally behave. Also, this strategy would imply that the human operator has full information about the state of the automation, which is often not the case. Accordingly there is an interest in the specific system properties that foster an appropriate reliance on automated systems and in the conditions that cause overreliance or disuse. Among a system's properties, its performance, especially its reliability, has gained special interest in this respect, evidenced also by its appreciation in theoretical models of reliance.

2.3.3.1 Operators' sensitivity to varying levels of automation reliability

Numerous studies have shown that operators are sensitive to automation reliability. In a laboratory study participants were presented with alarms of varying reliability (manipulated by the percentage of correct vs. false alarms) in a dual-task situation (Bliss, Gilson, & Deaton, 1995). Participants learned to match their response frequency to the probability of true alarms when there was no mean to validate the correctness of the alarms. Also in a laboratory environment, Getty, Swets, Pickett, and Gonthier (1995) varied the positive predictive value of a warning (i.e. the probability that a warning will truly indicate some specified dangerous situation) and examined participants' response latency to those warnings. A bonus scheme was introduced for participants earning bonuses for accurate performance in a background tracking task and for rapid responses to true alarms. Participants were found to generally behave in a rational manner by responding slowly to low positive predictive value warnings and quickly to high positive predictive value warnings to maximise the net gain from the benefits and penalties imposed. In a driving simulator study, Bliss and Acton (2003) varied the reliability of rear-end collision alarms (percentage of true vs. false alarms). Drivers had to respond to true alarms by performing preventive actions in order to avoid collisions with an approaching vehicle. Bliss and Acton found that drivers' response frequency increased as the alarm reliability increased. However, drivers that were presented with 100% reliable alarms were found to collide with significantly more cars than drivers that were presented with 50% reliable alarms. The researchers reported that in progress of the experiments, some participants in the 75% and 100%

alarm reliability group did not rely on the rear-view mirror to determine the direction of the approaching car (either from the left or from the right) but seemed to routinely perform some slighter swerving reaction which may have resulted in a higher number of collisions. The latter finding can be interpreted as a sign for drivers' overreliance on the highly reliable warning system (Bliss & Acton, 2003). In another driving simulator study, Maltz and Shinar (2007) also varied the reliability of a forward collision warning system, but this time by introducing false alarms and misses. Although the system was effective in reducing the proportion of time drivers spent with short headways when they were distracted by a loading task, Maltz and Shinar found that when time headway was less than 1 s and no alarm was given, drivers in the low reliability condition slowed down for a higher fraction of time than drivers in the medium and high reliability conditions and than drivers in the control condition. This was attributed to the increased awareness for short headways of the drivers in the low reliability condition as they had learned to rely less on the warning system. Yeh and Wickens (2001) manipulated the reliability of an intelligent target cuing aid in a target discrimination task. They found that although 100% reliable target cuing significantly improved the detection accuracy of low-salience targets, it significantly impaired detection performance for unexpected, but high-priority targets (the unexpected targets were never cued directly, but presented within 15° of an expected target). This effect was absent when the target cuing was 75% reliable. Apparently, the less reliable aid reduced attentional tunnelling and participants were more likely to examine the raw data.

2.3.3.2 Analysis of operators' reliance on binary warning systems by means of signal detection theory

Operators' reliance on binary warning systems or decision aids can be analysed in the framework of Signal Detection Theory (SDT; Green & Swets, 1966). Warning systems are characterised by presenting "at any moment in time, one of two or more messages (inactivity of the warning is also considered to be a message), based on input from sensors, people, or computational devices. At least one (but not all) of the messages alerts the user about a potential hazard that may require an intervention or closer monitoring" (Meyer, 2004, p. 196). In terms of signal detection theory, the use of a warning system or decision aid can be considered as the combined performance of two detectors: the operator and the warning system (Sorkin & Woods, 1985). Both observe some partially correlated noisy input channel and make binary decisions about the state of the monitored system (normal, abnormal). Both detectors can be described by two parameters: a sensitivity parameter (d') which specifies how effectively the monitor can discriminate signal-plus-noise events from noise-alone events on the channel, and a response criterion parameter (or cut-off point) which determines how much evidence is needed that an event is assigned to one distribution rather than the other. The output of the automated monitor (warning system) serves as input for the operator, who may use the decision of the warning system to combine it with other available information and respond according to the results of this combination. The differential adjustment of the operator's response criterion to the different outputs of the warning system can serve as a measure of his or her reliance. When a warning is valid, the probability of an abnormal state is larger when a warning is given than when no warning is given. For a non-valid system the probabilities for malfunctions are the same whether or not there is a warning. When an operator uses the same response criterion irrespective of whether a warning is given or not, he or she obviously ignores the warning. The use of different criteria however is evidence that the operator considers the warning and assigns different weights to the output of the warning system (Meyer, 2001). Also, operators' responses may be related with different consequences given the outputs of the warning system. For instance, the negative consequences from not detecting a malfunction are often more severe when a warning was given than when it was not. An expected value analysis of the payoffs of operators' responses given the outputs a warning system suggests that there are two different responses to a warning system: *compliance* and *reliance*[1] (Meyer, 2004). *Compliance* refers to the tendency to perform an action when the warning system instructs the operator to perform this action, whereas *reliance* refers to the tendency to refrain from performing an action when the warning system does not indicate that it is necessary. Compliance and reliance may be differently affected by automation false alarms and misses (Dixon & Wickens, 2006; Dixon, Wickens, & McCarley, 2007; Parasuraman & Wickens, 2008; Wickens & Colcombe, 2007).

In the aforementioned study, Yeh and Wickens (2001) also employed SDT in order to assess participants' reliance on the target cuing aid. This analysis revealed a shift in the response criterion when cuing was available without a corresponding increase in sensitivity, so that participants were riskier in their responses and likely to overlook errors in the cuing aid, resulting in a higher false alarm rate. When the target cuing aid was less reliable, participants adjusted their response criterion to a more conservative value and their sensitivity increased, but still they were guided by the advices of the cuing aid, as evidenced by a lower and riskier setting of the response criterion in the cued compared to the uncued condition. Participants reliance on the cuing aid despite its low reliability could be attributed to the fact that the target discrimination task was difficult (Yeh & Wickens, 2001). Similar results were found in a study by Maltz and Meyer (2001) who assessed the use of a cuing aid in a detection task. Participants tended to ignore the non-valid and low-validity cues and to rely only on highly valid cues, as evidenced by their response criterion setting for the different outputs of the cuing aid (alarm vs. no alarm). However, in all conditions (especially in those presented with non-valid and low-validity cues) participants adopted response criteria that were less cautious than the optimal criterion setting. Interestingly, results also revealed that the mere presence of the cuing aid affected participants' general tendency to take risks, even when the cues were non-valid and not directly used. This was established by comparing the participants' criterion settings in the non-valid warning conditions with those of the control condition. Thus the mere existence of a warning system may create the impression that the system relieves operators of part of the responsibility for the task at hand, allowing operators to take greater risks. Results of a

---

[1] The term reliance in this context refers to one aspect of the superordinate concept of reliance as it is used in the present work. In order to avoid confusions, authors who distinguish between compliance and reliance refer to the superordinate concept as operator *dependence*.

previous study (Meyer, 2001) showed a strong initial reliance on the binary cue based on its validity, followed by a gradual shift of reliance towards the optimal setting. Participants in the non-valid warning condition learned to respond more cautiously when the warning indicated normal functioning, but did not alter their response strategies when the warning indicated a malfunction, yielding evidence for the two different types of reliance. Similarly to the results of Maltz and Meyer (2001), participants continued to rely to some extent on the non-valid cues, but less than on valid cues.

2.3.3.3 Automation bias: Omission and commission errors

Similar to Meyer's (2004) distinction between reliance and compliance, Mosier and Skitka (1996) differentiate between errors of *omission* and errors of *commission* resulting from inappropriate reliance on automated decision aids. Errors of *omission* occur when operators fail to notice system irregularities or malfunctions because an automated aid failed to detect or to indicate them, comparable to Meyer's notion of reliance. Errors of *commission* occur when operators incorrectly follow the recommendations of an automated decision aid because they do not verify it against other available information, similar to Meyer's notion of compliance. According to Mosier and Skitka (1996), both types of errors result from operators' "tendency to use automated cues as a heuristic replacement for vigilant information seeking and processing" (p. 205), which the researchers termed *'automation bias'*. Skitka, Mosier, and Burdick (1999) found evidence for both types of errors in a low-fidelity simulated flight task that required participants to monitor a set of subsystems. Participants in the automated condition were assisted by an automated monitoring aid that notified them of critical events and recommended specific course of actions. Participants were told that the aid was highly, but not perfectly reliable, and that other system indices (the "raw" information) were however 100% valid. On 6 out of 100 events the automated aid failed to announce a critical system state, representing an opportunity for omission errors. Comparison of detection performance on these 6 events revealed that participants assisted by the automated decision aid missed significantly more events (41%) than participants in the control condition that were not assisted by such an aid (3% misses). The decision aid also incorrectly indicated a critical system state for 6 out of 100 events (false alarms). Participants were shown to incorrectly respond on the average to 65% of these events (commission error rate). A subsequent study yielded evidence that the automation bias (and hence, omission and commission error rates) could be mitigated when participants perceived themselves as being accountable for overall performance and accuracy (Skitka et al., 2000). Skitka et al. conclude further, that "errors of commission proved to be the result of cognitive vigilance decrements" (Skitka et al., 2000, p. 52). Interestingly, the researchers did not explicitly assess participants' vigilance. The results of the study rather suggest that participants altered their performance strategies according to the pay-off structure of the task which was induced by different instructions. Indeed, the analysis of omission and commission error rates may not give much insight into operators' reliance without considering the expected value of operators' responses (Meyer, 2004).

2.3.3.4 Summary

In summary, the above review of empirical studies suggests that operators are sensitive to varying levels of automation reliability. Studies on binary warning systems show that operators are able to differentiate among varying diagnostic values of a warning and adjust their response strategies according to that. Operators were shown to be more careful in responding to the outputs of a warning system when the reliability or the diagnostic value of the warnings was low compared to high. Nevertheless, almost all studies found evidence that operators' reliance on warning systems was not optimal in that they relied even to some extent on non-valid warnings. Evidence for this was found in terms of operators' riskier setting of the response criterion under the presence of a warning system compared to manual control. Thus, from a rational point of view, there is indeed an indication for operators' tendency to overrely on automation. In the framework of the reported studies this at least indicates that operators have the justifiable assumption that the presence of automation implies any rationale in terms of its usefulness or beneficial effect for performance and efficiency. On the other hand, there is some evidence that operators' task-related effort and their active task engagement decline the higher the reliability of the automation.

Aside from motivational processes, other factors, such as operators' attitudes towards automation and energetic processes are also expected to mediate the influence of automation reliability and of the level of automation on reliance. Empirical evidence regarding the mediating influence of these factors will be reviewed within the next chapters.

2.3.4 Operators' attitudes towards automation and their influence on reliance

2.3.4.1 Trust in automation

The calibration of trust

Substantial evidence suggests that trust in automation is a meaningful and useful concept to understand operators' reliance on automation. Trust can help people to overcome complexity and uncertainty by acting as a social decision heuristic in guiding reliance. Trust enables operators in the first place to rely on an automated system when its exact properties are not known, especially at an initial use of a system. On the other hand, relying on heuristics in the decision to rely on an automated system may not be appropriate in any situational context. The demand for well-calibrated trust that matches the true capabilities of a system gains its importance in this point (Lee & Moray, 1994; Lee & See, 2004; Muir, 1987, 1994; Sheridan & Hennessy, 1984). Well-calibrated trust is characterised by high *resolution* and high *specificity* (Cohen, Parasuraman, & Freeman, 1999; Lee & See, 2004). High *resolution* means that trust differentiates well between varying levels of system capabilities, i.e. large changes in system capabilities map onto large changes in operators' trust (Cohen et al., 1999; Lee & See, 2004). *Specificity* refers to the degree to which trust differentiates between particular aspects, components or modes of a system (functional specificity) and to the temporary adjustments of trust according to the moment-by-moment fluctuations in system capabilities (temporal specificity). High *functional specificity* is given when trust reflects the performance or the capabilities of particular subfunctions or modes of a system rather than the performance of the entire system. High *temporal specificity* means that trust is sensitive to changes in situation or context that affect the capability of the automated system. For example, a Lane Departure Warning system may produce highly reliable warnings of an imminent unintentional departure from the driving lane on road segments where the lane markings are clearly identifiable, but not on road works sections with (for the system undistinguishable) overlaying temporary and original lane markings. High specificity of trust would mean that the driver is aware of this change in context and the accompanying reduced reliability of the system and (temporarily) adjusts his or her level of trust according to this limited system performance. The importance of transparent system design for making functional system limits visible for the driver was stressed in a number of works (Goodrich & Boer, 2003; Nilsson, 2005; RESPONSE 3 Code of practice for the design and evaluation of ADAS," 2006; Seppelt & Lee, 2007; Simon, 2005). Goodrich and Boer (2003) propagate a "model-based human-centred task automation" that requires the identification of driver's mental models for specific driving subtasks in order to use them as a template for automation: "If the limits of automation correspond to the limits of a subset of natural operator skills, then the limits of the automation are most likely to be perceived and detected by the operator" (p. 329). They demonstrate this approach for the design of an ACC system.

The basis of trust

Trust is built on information that informs the human operator about the ability of the system to achieve the operators' goals. The design of the human-machine interface plays an important role in the evolution and the calibration of trust as it determines how well this information is conveyed to the operator. According to Lee and See (building on the work of others; e.g., Lee & Moray, 1992; Muir, 1987; Muir, 1994), trust is based on three types of goal-related information regarding a system's *performance* (derived from a direct observation of system behaviour), *process* (based on an operator's understanding of the underlying mechanism), and *purpose* (based on the intended use of a system). Performance information describes *what* the automated system does and includes system characteristics such as reliability and predictability. Process information describes *how* the automated system operates, i.e. the degree to which the algorithms underlying system performance seem appropriate in a given situation to achieve the operators' goals. Purpose information refers to the aim of the automated system, i.e. *why* it was developed. People are more likely to accept systems that they regard as useful. Trust is not only affected by the availability of information regarding these three dimensions, but also by the inferences operators draw among them. Observation of a system's performance in different contexts or situations is likely to affect operators' beliefs about the system's internal mechanism guiding its behaviour (process information). Likewise, knowledge of the intended purpose of a system (e.g., as communicated by manufacturers) presumably influences operators' expectations regarding its performance and their assumptions about how it might operate. Inconsistencies between inferences drawn on the three dimensions as well as mismatches between inferences and observations of real system behaviour are expected to have negative effects on trust and hinder a good calibration of trust.

The influence of automation reliability on trust and reliance

A majority of research has focused on the influence of a system's performance on human trust and reliance. Substantial empirical evidence was found that trust in an automated system varies as a function of its reliability (De Vries et al., 2003; Lee & Moray, 1992, 1994; Lewandowsky et al., 2000; Metzger & Parasuraman, 2005; Moray, Inagaki, & Itoh, 2000; Muir & Moray, 1996). In many of these studies participants acted as supervisors of a simulated process control plant in which different subsystems could be operated under automatic or manual control. The reliability of automated subsystems was manipulated by introducing faults reducing overall system efficiency. Participants' reliance was measured by means of the proportion of time they chose manual vs. automatic control for operating subsystems prior and after the occurrence of faults. Participants' trust in the automatic subsystems was assessed via subjective rating scales. A general finding of these studies is that trust decreased in response to automation faults, and then recovered afterwards again. However, the recovery of trust from faults was not instantaneous and was slower than the recovery of performance, demonstrating an effect of inertia (Lee & Moray, 1992, 1994). Furthermore, trust seemed to develop steadily and slowly when the performance of a system was high, but was markedly affected by the occasional occurrence of a fault (Lewandowsky et al., 2000; Moray et al., 2000; Muir & Moray, 1996). Trust seems to be not only affected by the magnitude of a fault, but also by its predictability. A small unpredictable (variable) fault caused trust to decline and then to remain stable and low, whereas a constant, relatively large fault caused trust to decline to a lower initial level, but than to recover as participants found strategies to compensate for it (Muir & Moray, 1996). Muir and Moray also found that trust was greatly reduced by small control errors, but was increasingly less sensitive to larger control errors. Moray et al. (2000) found that trust was little affected when reliability was above 90%, but was increasingly strongly affected when reliability fell below this level.

The magnitude and the predictability of faults had also an effect on participants' monitoring of the subsystem under automatic control. In the study of Muir and Moray (1996) participants could monitor the accuracy of an automatic pump by typing requests for the current pump rate to be displayed on a control screen. The researchers found that operators checked the status of the automatic pump more frequently when it worked less reliable and when unpredictable (variable) errors occurred compared to constant ones. Moreover, Muir and Moray found a negative correlation between participants' monitoring behaviour and their ratings of trust in the automatic pump: The less they trusted it, the more intensely they monitored it and the more they trusted it, the less intensely they monitored it; supporting former predictions by Muir (1994) and Sheridan and Hennessy (1984).

Reliance on automation (measured by the percentage of time spent in automatic control) could not be consistently predicted by operators' trust. Whereas Muir and Moray (1996) found that operators' trust was a very good predictor of their reliance, others (De Vries et al., 2003; Lee & Moray, 1992, 1994; Lewandowsky et al., 2000) found that operators' reliance could be more appropriately predicted by the relationship between their trust in the automation and their self-confidence in manually controlling the system. Specifically, Lee and Moray (1994) found that participants' reliance could be predicted by a logit function of the difference between trust and self-confidence, whereby in general operators tend to rely more on automation when their trust in the automation exceeds their self-confidence, and tend to engage in manual control when trust was lower than their self-confidence. Furthermore, by fitting a time series model to the data, Lee and Moray (1994) could show that participants' reliance depended not only upon the difference between trust and self-confidence, but also on previous reliance on automation (reliance on automation in one trial generally led to reliance on automation in the next trial) and upon individual biases (some participants consistently preferred manual over automatic control regardless of their ratings of trust and self-confidence, and vice versa). Recently, Gao and Lee (2006) could replicate these findings by using extended decision field theory to model operator reliance on automation.

Summary

In summary, there is substantial empirical evidence for the mediating influence of trust in automation on operators' reliance. Trust influences reliance as an attitude which is based on operators' perception of the automation's ability to serve their task-related goals, e.g. to maintain acceptable standards of performance and system efficiency. Thus, the reliability of an automated system has a significant influence on operators' trust. Operators' trust in automation was found to increase as a function of increasing reliability. In turn, operators seem to be increasingly willing to allocate control to an automated system and to reduce their efforts for critically monitoring its actions the more they trust it.

Up to now, there are no empirical studies on the effects of different levels of automation on operators' trust. Theoretical frameworks on trust and reliance make no assumptions about such a relationship. It is however assumed that unreliability has a greater negative impact on operators' trust and reliance the higher the level of automation (Gao & Lee, 2006; Lee & See, 2004; Parasuraman et al., 2000). That is because even a low level of reliance on high-level automation is necessarily associated with the allocation of a substantial amount of control to an automated system. That means that in order to be able to observe the behaviour of high-level automation, operators have to rely to some extent on it. With low-level automation, e.g. with a system that provides operators with additional information about the environment, operators have generally still access to the "raw" data and are able to observe and verify the system's behaviour without being forced to rely on it. The impossibility to observe the behaviour of high-level automation without allocating a substantial amount of control to it may lead to nonlinear patterns of trust and reliance following automation malfunctions. There is some empirical evidence that unreliability is associated with greater performance costs when the level of automation is high compared to low (e.g., Galster & Parasuraman, 2004; Rovira, McGarry, & Parasuraman, 2007)

At present, it is not completely clear how regularly operators update their level of trust in an automated system. The regular updating of trust is however necessary to ensure its proper calibration in light of the situational context and the momentary performance of automation. Researchers have propagated the appropriateness of trust as a prerequisite for operators' appropriate reliance on automation (Lee & See, 2004). Not all studies have however found a linear relationship between trust and reliance (e.g., Bagheri & Jamieson, 2004; Rovira et al., 2007; Wiegmann, Rich, & Zhang, 2001). It may be possible that trust is more sensitive to momentary fluctuations in automations' performance during the first time of interaction with automation, while with growing system experience, operators may develop a rather stable level of trust which may be less predictive for momentary changes in reliance. Theoretical frameworks assume that reliance is influenced by also by other variables such as operators' energetic state and motivational processes which will be reviewed in the next chapters.

2.3.4.2 Complacency

A high level of trust or "blind trust" in highly reliable automation was also hypothesised as one possible cause for operators' complacency, as reported in a study by Parasuraman et al. (1993). Complacency is defined as "self-satisfaction which may result in non-vigilance based on an unjustified assumption of satisfactory system state" (Parasuraman et al., 1993, p. 2, quoting NASA definition). Complacency was cited as a major cause for the decline in operators' detection performance for system malfunctions when they monitored an automation in detecting those malfunctions compared to when they operated a system manually.

Operators' impaired detection of automation failures as an indicator for complacency: The Parasuraman et al. (1993) study

In the study of Parasuraman et al. (1993) participants had to perform three tasks (monitoring, tracking, fuel-management) of a flight-management simulation battery. The monitoring task required them to monitor a set of gauges to ensure that none of them went out of range. Normally, an automation routine detected and corrected those malfunctions. In some occasions however, the automation routine failed to detect gauges malfunctions. Participants were required to detect these automation failures and to reset the gauges manually by pressing a key. Parasuraman et al. manipulated the reliability of the automation routine (either low or high) and the consistency of the automation's reliability (either constant-high or variable) yielding four experimental groups. For the two constant-reliability groups the reliability of the automation routine remained either high or low throughout the course of the experiment. For the two variable-reliability groups the percentage of the malfunctions detected by the automation routine alternated (from low to high or vice versa) from block to block. Within the last two blocks in each of the four conditions the automation routine was completely disabled representing zero reliability. Parasuraman et al. found that in a 10min training period where participants performed all of the three tasks manually, they detected 72% of the gauges malfunctions. When the monitoring task was automated afterwards, participants detected only 32% of the malfunctions. The level of reliability (low vs. high) did not significantly affect detection rates for automation failures. However, there was a marked effect for the consistency of automation reliability: Participants in the variable-reliability group detected significantly more automation failures than participants in the constant-reliability group (see also Singh, Molloy, & Parasuraman, 1997), which Parasuraman et al. attributed to a more sceptical attitude of the variable-reliability group regarding the capabilities of the automation routine which in turn promoted higher vigilance. The 'complacency effect' was found to arise only after 20 min, and was only apparent under multitask conditions and not when monitoring the automation for failures was the only task, as was demonstrated in a subsequent experiment.

Indeed, the latter fact may also argue for the view that participants had adopted a rational strategy given the consequences of non-detection of automation malfunctions (there were no negative consequences since the faulty condition was reset after 10s without intervention), the low frequency of automation failures (and the associated difficulty of participants to learn the probability of the occurrence of critical events) and their responsibility for the other two task. As Moray and Inagaki (2000) convincingly argued, complacency can only be measured if an optimal monitoring strategy is defined that considers the pay-off-structure of a task: "Since even optimal sampling can not detect all signals, research on monitoring and complacency should not use detection but sampling as a measure for the quality of attention" (Moray & Inagaki, 2000, p. 362).

Operators' monitoring of automation as an indicator for complacency and the nature of the complacency effect

Bagheri and Jamieson (2004) replicated the Parasuraman et al. (1993) study and also assessed participants' monitoring behaviour, their subjective trust in the automation routine, and their self-confidence. The researchers found that the detection performance of the constant-high reliability group was significantly poorer than the detection performance of the variable-reliability groups (replicating the finding of Parasuraman et. al), but also than the detection performance of the constant-low reliability group (unlike the results of Parasuraman et al.). Also, the detection performance of the constant-high reliability group was significantly poorer than the performance of the variable-reliability group in those blocks where the reliability of the automation routine was equally high. Interestingly, Bagheri and Jamieson also found that participants in the constant-high reliability group monitored the gauges display less frequently in the first blocks of the experiment, with the mean time between fixations gradually increasing from an initially comparable level for the constant-high reliability group, but than decreasing in the second half of the experiment and converging those of the other three groups. Even though this finding in principal corroborates the interpretation of the results in light of a complacency effect, it contradicts the nature of complacency as it is assumed to develop gradually over time when operators are faced with highly reliable automation. The four experimental groups did not significantly differ in their trust ratings; however, trust was negatively correlated with the detection performance for automation failures and with the time between subsequent checks of the gauges status (monitoring). Participants' self-confidence was lowest in the constant-high reliable group, and highest in the constant-low variable group. Altogether, Bagheri and Jamieson argue that this pattern of results does not support the attribution of the constant high-reliability group's poor detection performance to complacency. Furthermore, Bagheri and Jamieson tested the efficiency of participants' monitoring strategy against an optimal sampling model and yielded evidence that the monitoring strategy of the constant-high reliability group was more 'expensive' than an optimal sampling strategy and more expensive than the sampling strategy of any of the other experimental groups. However, the constant-high reliability group's sampling strategy became more efficient over time, which contradicts the evolution of complacency.

In a recent study on complacency, Bahner et al. (2008) directly assessed participants' information sampling behaviour in order to investigate to what extent the recommendations of an automated decision aid were cross-checked by the participants against other available information in a process control simulation. The automated decision aid had two functions: First, it generated an alarm as soon as detecting a fault in any subsystem; and second, it displayed both a fault diagnosis and a supposed sequence of actions for effective fault management. The warning function was completely reliable. However, the automated aid one time provided a false fault diagnosis and two times failed to provide any fault diagnosis at all towards the end of the experiment. During the trainings session with the automated aid, only half of the participants actually experienced failures of the automated aid to provide correct fault diagnoses. The other half of the participants were just informed that the fault diagnoses provided by the automated aid might be incorrect. The results show that both groups of participants were complacent to some extent in that they sampled less information than needed to fully verify the automated aid's diagnoses before starting to intervene in the system. However, complacency was more pronounced for participants who were just informed that the aid might provide false diagnoses than for participants who actually experienced failures of the automated aid during training. Participants who actually experienced false diagnoses invested considerably more effort in verifying the automated aid's fault diagnoses. Specifically, they were found to sample a higher proportion of relevant information and needed more time for fault management than participants who were just informed about possible failures of the automated aid. The two groups did not differ with respect to their detection rate of the automated aid's wrong diagnosis occurring towards the end of the experiment. However, there were marked differences between participants who detected the false diagnosis and participants who did not and followed the incorrect fault management recommendations of the automated aid. Participants who did not detect the automated aid's false diagnosis were found to having had spent considerably less effort in verifying the decision aid's recommendations during previous trials. Specifically, they had sampled less information and a smaller portion of relevant information during previous trials and needed only half of the time for fault identification than participants who detected the false diagnosis. Furthermore, participants who did not detect the aid's false diagnosis showed a significantly stronger decline in information sampling behaviour over time than participants who correctly identified the aid's false diagnosis. Unfortunately, Bahner et al. (2008) did not collect subjective measures of mental workload and of participants' trust in the decision aid.

Summary

In summary, it appears that operators' monitoring behaviour is closely linked to the performance of an automated system. Operators' detection performance of critical system states or automation malfunctions was found to severely deteriorate (in terms of accuracy and response times) when they experienced the automation to be highly reliable. There is converging empirical evidence that these effects at least partially result from operators' diminished efforts in monitoring the automation. Monitoring an automated system for its proper functioning can be considered effortful and resource demanding, especially when the system automates cognitive functions such as human decision making. Furthermore, complete verification of an automated systems' behaviour also involves performance costs such as an increase in response times to system faults due to cross-checking of other information (Bahner et al., 2008; Ezer, Fisk, & Rogers, 2008). Thus, human operators seem to have a tendency to rely on automation unless they have substantial doubts in its proper functioning. The mere knowledge that an automated system may fail does apparently not cause operators to invest more mental effort in monitoring and verifying the automations behaviour.

Alternatively to motivational explanations, Farrell and Lewandowsky (2000) provided a learning-based account for the complacency effect. The basic assumption in their computational model is that the difficulties operators have in detecting and responding to automation failures arise from response competitions and interference of different cognitive operations acquired during manual control and under automation. Specifically, "under automation, people learn to *withhold* a response, whereas under manual control, people learn to *execute* a task-appropriate response. Performance suffers when a different cognitive operation is suddenly required – for example, when an automation failure calls for a compensatory response. Critically, on this account, complacency does not arise from automation per se; instead, the skill to withhold a response that was acquired during automation cannot be usefully reinstantiated when the need for a manual response arises" (Farrell & Lewandowsky, 2000, p. 397).

Although it is generally assumed that complacency and its effects on reliance are caused by a high or even excessive level of trust in automation capabilities, empirical studies do not provide strong support for this hypothesis. It appears that compared to trust, external task demands have a stronger influence on operators' effort management strategies and on their monitoring behaviour. In line with this, the complacency effect was only found when operators were engaged in multiple tasks and not when monitoring the automation was the only task. Reliance on automation and a neglect of automation monitoring when the automation appears highly reliable may therefore reflect an effective workload and attention management strategy. An argument against this assumption seems to be that operators who relied to a greater extent on automation did often show no better performance in concurrent tasks, although they should have had more spare mental capacity. Two explanations may account for this finding. First, the secondary tasks employed may not have been sensitive enough for eliciting changes in participants' attention allocation strategies. Second, participants may not try to maximise performance through complete effort expenditure, but rather try to yield an acceptable overall performance across concurrent tasks while conserving mental effort (cp. Wickens & Dixon, 2007).

Trust in automation and an individual predisposition for complacency however seem to account for individual differences in monitoring behaviour and automation reliance. For example, in the replication study of Bagheri and Jamieson (2004) the inter-group differences in detection performance of automation failures could not be explained by differences in participants' trust. However, on an individual level, participants' trust was negatively correlated with detection performance of automation failures and with their sampling behaviour. Likewise, Bahner et al. (2008) found that participants who did not detect an automated decision aid's failure had previously invested considerably less effort in verification of the automated aid's decisions, and showed a larger decline in information sampling behaviour over time than participants who did detect the automated aid's failure. Prinzel et al. (2005) provided evidence for the moderating effect of participants' complacency potential on their detection performance of automation failures under constant high and variable (alternating low and high) reliability conditions. Prinzel et al. found that the decline in detection rate of automation failures under constant high reliability conditions was significantly more pronounced for participants who were diagnosed as high in complacency potential. Participants with a high complacency potential performed significantly worse than participants with low complacency potential in a concurrent tracking task in the variable reliability condition which was presupposed to induce higher demands for automation monitoring. Likewise, high complacency potential participants reported a significant larger increase in subjective workload in that condition. On the contrary, they reported less mental workload than low complacency potential participants in the constant high reliability condition, which was attributed to their neglect of monitoring the automation that in turn resulted in worse detection performance of automation failures.

To conclude, empirical evidence suggests that operators invest less effort in automation monitoring when they perceive the automation as performing highly reliable. There are furthermore individual differences in operators' tendency to adopt such effort conservation strategies as an effective mean to deal with high workload. It has to be shown whether the level of automation has a similar effect on operators' reliance; that is whether systems that automate more aspects of a task similarly reduce operators' efforts to stay actively engaged in a task and to monitor the automated system's actions.

2.3.5 Energetic processes and their influence on reliance: Vigilance, arousal and mental workload

Reliance on automation is assumed to be not only influenced by operators' attitudes towards automation (such as trust), but also by their energetic state which can be described by concepts such as arousal, mental workload, and vigilance. The interest in those energetic concepts in mediating reliance has emerged from the changes in task demands due to automation. A critical requirement of supervisory control is the efficient monitoring of automated system functioning, which is usually operationally defined by operators' detection rate of system malfunctions and/or failures. Monitoring of proper system functioning is clearly a vigilance task (Molloy & Parasuraman, 1996; Parasuraman, 1987; Scallen & Hancock, 2001; Warm, Dember, & Hancock, 1996; Warm, Parasuraman, & Matthews, 2008). *Vigilance*, or *sustained attention*, refers to "the ability of observers to maintain their focus of attention and to remain alert to stimuli for prolonged periods of time" (Warm et al., 1996, p. 183). Typical vigilance tasks require observers to monitor displays over extended periods of time for the occasional occurrence of critical events or signals (Davies & Parasuraman, 1982). Thus, a vigilant state is a precondition for efficient system monitoring. Conditions or factors that impair vigilance also impair the operator's ability for system monitoring, and thus may favour reliance on an automated system in the sense of increasingly handing over control to a system (coupled with the tendency to neglect checking the appropriateness of such behaviour).

There are at least two arguments why automation may impair vigilance, and thus, monitoring. The first argument lies in the vigilance decrement itself. The *vigilance decrement* refers to the decline in detection performance over time and is usually manifested as a drop in detection accuracy and/or an increase in response times. Thus, over prolonged periods of system monitoring, the operator's attentional resources become exhausted, and this depletion of resources impairs the capability for subsequent vigilant monitoring. As a consequence, operators' (uncritical) reliance on an automated system may increase.

The second argument refers to the potential negative effect automation may have on mental workload. Especially high-level automation may relieve the operator from a substantial amount of the control processes involved in tasks, and thus may cause *underload* and a low level of *arousal* which subsequently impair vigilance, and in turn likely foster reliance. A third argumentation line refers to the impact of operator attitudes (such as *trust* and *complacency*) on system monitoring. There is empirical evidence that high levels of trust and complacency lead to a neglect of monitoring. However, at present it is unclear whether this monitoring deficiency is moderated by a decline in sustained attention, i.e. whether trust and complacency have a direct impact on the operator's energetic state and subsequently impair his or her monitoring efficiency.

2.3.5.1 Increasing reliance due to an exhaustion of attentional resources

There is converging empirical evidence that the vigilance decrement indeed reflects the depletion of information-processing resources that cannot be replenished in the time available (Davies & Parasuraman, 1982; Grier et al., 2003; Helton et al., 2005; Hitchcock et al., 2003; Matthews & Davies, 1998; Warm & Dember, 1998; Warm et al., 1996; Warm, Matthews, & Finomore, 2008; Warm & Parasuraman, 2007; Warm, Parasuraman et al., 2008). For example, Hitchcock et al. (2003) used transcranial Doppler sonography (TCD) to examine the influence of automation cues of varying reliability on vigilance performance in a 40-min simulated air traffic control task. It was hypothesised that providing participants with reliable cues as to the imminent arrival of critical signals would improve vigilance performance and attenuate the vigilance decrement, especially under low salience conditions. Participants were informed about the cue reliability before the experiment started. Indeed, Hitchcock et al. found that detection performance of the control group (without advanced cueing information) declined most over time, followed in order by the 40% reliable, 80% reliable and 100% reliable cue groups The percentage of correct detections remained almost stable for the 100% reliable cue group. The changes in detection performance were mirrored by a temporal decline in cerebral flow velocity in the right hemisphere, yielding the highest blood flow in the 100% reliable cue group, followed in order by the 80% reliable, 40% reliable, and no-cue groups. This decline in cerebral blood flow velocity was not found in a group of participants who only passively gazed at the display where the experimental task was presented. Thus, the changes in blood flow velocity appear to reflect a decline in the utilisation of information-processing resources in sustained attention and fit well with an attentional resource model of vigilance (Hitchcock et al., 2003; Warm, Matthews et al., 2008; Warm & Parasuraman, 2007; Warm, Parasuraman et al., 2008).

Furthermore, in the Hitchcock et al. (2003) study, the changes in detection performance were also paralleled by a decline in energetic arousal measured by a scale for that dimension on the Dundee Stress State Questionnaire (DSSQ; Matthews et al., 1999). Energetic arousal scores were similar to each other pre-vigil, but were different post-vigil, with scores dropping to the lowest level in the no-cue condition, followed in order by the 40% reliable, 80% reliable, and 100% reliable cue group. This finding indicates that “performance related changes in blood flow and energetic arousal may share some common mechanism” (Hitchcock et al., 2003, p. 107). Thus, arousal may be viewed as the agent that is responsible for the production of resources (Hitchcock et al., 2003; Humphreys & Revelle, 1984; Matthews et al., 1990; Warm et al., 1996).

Other research has shown that vigilance tasks are very resource demanding and impose a substantial workload on human operators. More specifically, Warm et al. (1996) found that the vigilance decrement was accompanied by a linear increase in overall workload over time. When mental workload is assessed subjectively by means of the NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988), vigilance tasks produced a consistent workload signature among subscales with mental demands and frustration being the primary components of workload.

In summary, there is consistent empirical evidence that vigilance tasks such as system monitoring are highly demanding and deplete attentional capacity and information-processing resources over time. The result is the well-known vigilance decrement. It is however not clear at present whether this decrease in attentional resources also reduces the effort operators invest in subsequent automation monitoring and thereby fosters an uncritical reliance on automation. Theories on mental effort regulation would predict that after prolonged periods of mental effort investment (such as during vigilance tasks) operators will shift to task strategies where mental effort is minimised (Fairclough, 2001; Hockey, 1997). Thus, it can be hypothesised that over time operators are likely to invest fewer resources in system monitoring and to be more prone to uncritically rely on an automated system, especially when it is highly reliable. Bahner, Hüper, and Manzey (2008) found some evidence for this by assessing participants' information sampling behaviour when interacting with an automated decision aid in a process control simulation. Information sampling (measured by the total number of sampled information units per second) significantly decreased over experimental blocks.

2.3.5.2 Increasing reliance due to underload

Concerns have been raised that automation may not only increase mental workload by assigning the operator a monitoring role, but may also leave the operator with too little things to do and thereby instead induce underload (Byrne & Parasuraman, 1996; Carsten & Nilsson, 2001; Desmond & Hancock, 2001; Desmond, Hancock, & Monette, 1998; Hargutt & Krüger, 2001; Matthews, 2002; Parasuraman & Hancock, 2001; M. S. Young & Stanton, 2002a, 2002b). Theoretical frameworks assume that there is an optimal state for the human operator in a given task environment. Both underload and overload represent non-optimal conditions that induce stress and strain and require adaptive responses from the operator (e.g., Hancock & Warm, 1989). Underload is often conceptually linked with a low level of activation or arousal which is assumed to impair the mobilisation of attentional processing resources required for vigilant monitoring of an automated system's behaviour. Thus, underload and a low level of energetic arousal may favour reliance on an automated system by impairing the operator's capabilities for vigilant system monitoring.

There are indeed a number of studies that show that automation can reduce mental workload (e.g., De Waard et al., 1999; Harris, Hancock, Arthur, & Caird, 1995; Hoedemaeker & Brookhuis, 1998; Ma & Kaber, 2005; Masalonis, Duley, & Parasuraman, 1999; Stanton & Young, 2005; M. S. Young & Stanton, 1997). However, studies that evaluate the effect of increasing levels of automation on both operators' mental workload *and* on energetic arousal are sparse. The existing empirical studies suggest that there seems to be indeed an adverse effect of increasing automation (partly operationally defined as automation reliability) on arousal.

In a driving simulator study, De Waard et al. (1999) tested different conditions of automated and manual driving. In the two automated driving conditions, drivers drove in a platoon of cars. Lateral and longitudinal control was automated, but time headway to the car in front was either long (1 s) or short (0.25 s). Drivers' self reported mental effort and activation level were significantly lower during automated driving than during manual driving. Interestingly, the long time-headway condition again differed significantly from the short time-headway condition. The long time-headway condition was perceived as being significantly less risky, which was accompanied by significantly lower ratings of mental effort and activation.

In a laboratory study, Buld and Krüger (2002) varied the reliability of an automation routine in order to simulate increasing levels of automation (with high reliability corresponding to a high level of automation). The automation routine assigned balls of either red or green colour in one of two boxes with corresponding colour. The participants' task was to press a key as soon as they had realised that the automation routine would assign a ball to a box with a different colour (e.g., assign a red ball to the green box). Buld and Krüger (2002) found that reaction times for detecting automation failures linearly increased with reliability, i.e. the higher the reliability of the automation routine, the longer it took participants to respond. This increase in response time was mirrored by an increase in EEG alpha power density, but not by an increase in subjective ratings of sleepiness. In a subsequent driving simulator study the researchers investigated the effect of varying reliability of an Adaptive Cruise Control (ACC) system on subjective and physiological measures of arousal. Participants performed four 45-min drives: one manual drive and three drives with an either 95 % reliable, 75% reliable and 50% reliable ACC system. Results showed that participants felt increasingly bored, tired and less attentive the higher the reliability of the ACC system. There was no time-on-task effect for subjective measures of arousal. A time-on-task effect was however found for physiological measures, with EEG alpha power density reaching a similar high level in all conditions at the end of the study. At the beginning of the study, differences in alpha power density were found for the four conditions, with the highest alpha power density (corresponding to the lowest level of activation) during the 95% reliable condition, followed in order by the 75% reliable, 50% reliable and manual driving condition.

The aforementioned studies suggest that highly reliable automation that minimises the necessity for manual interventions may indeed lower operators' activation level. A low level of arousal has also been linked to impairments in sustained attention. Studies by Matthews and colleagues (Matthews & Davies, 1998; Matthews et al., 1990) suggest that energetic arousal is related to the availability of a resource required for sustained attention. Moreover, the facilitative effects of energetic arousal on performance were found to be largely linear rather than curvilinear as predicted by the Yerkes-Dodson Law (Yerkes & Dodson, 1908).

Other more indirect evidence for the negative effect of low arousal on sustained attention comes from studies that found that drivers used strategies to increase driving task demands in order to combat boredom, monotony and to diminish alertness. Such strategies include particularly adjustments in longitudinal control, such as alternating speed (Tejero & Chóliz, 2002) or increasing speed (Hargutt & Krüger, 2001) with prolonged duration of driving.

In conclusion, there seems to be a clear potential for automation to lower mental workload. There is also evidence that this lowered mental workload is associated with reduced activation and arousal, which is seen as being detrimental for performance in tasks that require sustained attention, such as system monitoring. Still, there is some lack in empirical studies that systematically varied the level of automation in order to investigate its effects on operators' mental workload, energetic arousal and reliance. Studies with varying levels of system reliability suggest that more capable systems requiring fewer manual interventions (in case of system failures) lead to lower levels of arousal and impair monitoring efficiency. Although yet not empirically tested (apart from some evidence in a study by Bahner et al., 2008) it can be assumed that due to the reduced capability for vigilant monitoring operators' tendency to uncritically rely on an automated system may increase.

2.3.5.3 Summary

Monitoring of automation is a resource demanding vigilance task. Thus, it can be assumed that operators' tendency to rely on an automated systems increases to the extent that their ability for vigilant system monitoring decreases. Automation may impair vigilance through two mechanisms: First, through resource depletion due to the added cognitive demands associated with automation monitoring; and second, through a low level of arousal caused by a reduction of workload due to automation. There is a lack of studies that systematically investigated the effect of different levels of automation on operators' vigilance, energetic arousal, monitoring behaviour and reliance. Buld and Krüger (2002) found that participants' activation level measured by EEG parameters decreased to the extent the reliability of an automated system increased. It can be hypothesised that high level automation has a similar effect on arousal and vigilance because it likewise reduces the need for operators to actively intervene in a task.

2.3.6 Motivational processes and their influence on reliance: Mental effort regulation

Reliance on automation is assumed to be also influenced by motivational processes and strategic decisions of the operators about the investment of mental effort. Current theoretical approaches assume that mental effort regulation is a cognitive mechanism by which operators strategically allocate processing resources in order to adapt to changing task demands (Fairclough, 2001; Hockey, 1997). Fairclough (2001) refers to two principle concurrent goals that operators have to prioritise according to the actual task demands, their performance, and the current state of their energetic resources: The first can be conceptually described as "perform well", and the second as "be comfortable". Prioritising the principle goal "perform well" over the goal "be comfortable" results in effort investment, whereas prioritising "be comfortable" over "perform well" results in strategies of effort conservation. Inherent in the view is the assumption of a limited pool of processing resources (Kahnemann, 1973). If task demands increase to such an extent that maximal effort expenditure is not sufficient to maintain performance, individuals react by changing their task goals in order to lower the effort required – either by lowering performance standards or by switching to low-effort strategies (Fairclough, 2001; Hockey, 1997; Parasuraman & Hancock, 2001). The current task goal is based on a utility judgement upon the costs and benefits associated with different performance strategies in light of the actual task demands and the effort investment they require (Hockey, 1997).

Theoretical models of reliance on automation assume that such a utility judgement influences operators' decision to rely on a system (Dzindolet et al., 2002; Lee & See, 2004). The utility judgement is based on the perceived performance of a system as well as on the perceived costs and benefits associated with different degrees of reliance (e.g., consequences associated with a system failure). System monitoring and the maintenance of active task engagement under partly automated control are considered very resource demanding (see chapter 2.3.5.1). This holds especially for higher levels of automation because the "raw" information on which an automated system decides and acts may no longer be accessible for operators. Thus, reliance on an automated system (usually associated with less efficient system monitoring) may appear as a useful effort conservation strategy particularly when the automated system works highly reliable.

2.3.6.1 Evidence for mental effort regulation in studies on operators' reliance on automation

Mental effort regulation was assumed to play a role for example in studies on complacency (Parasuraman et al., 1993). In the Parasuraman et al. study, detection performance of automation failures was only impaired under multitask conditions, but not when monitoring the automation for failures was the only task. Although participants' monitoring behaviour was not assessed directly, it may have been appropriate for them to put less effort in system monitoring and to allocate more resources to the other two tasks, especially because non-detection of automation failures did not have any negative consequences (see Moray & Inagaki, 2000). A counter-argument for that explanation seems to be that automation reliability did not play a role in that high automation reliability should have had led participants adopt the effort conservation strategy to a greater extent, leading to a worse detection performance of automation failures. However, in replication studies detection rates for automation failures varied inversely with automation reliability as expected, yielding a significantly worse detection performance when automation reliability was high compared to low (Bagheri & Jamieson, 2004; Bailey & Scerbo, 2007). Bagheri and Jamieson (2004) furthermore found that participants in the high automation reliability condition monitored the display of the automated subtask significantly less often, however, especially at the beginning of the experiment. Similarly, Bahner et al. (2008) found that when participants actually experienced false diagnoses of an automated decision aid before the experiment started they put more effort in sampling the relevant system parameters in order to verify the decision aid's diagnoses than when they were just informed about the possibility that the automated decision aid may err. Skitka et al. (2000) manipulated the pay-offs associated with system monitoring by instructing participants to prioritise detection accuracy of an automated decision aid's malfunctions over speed of responding and performance in a concurrent manual tracking task. Participants that were made accountable for their monitoring accuracy were found to verify to a greater extent the decisions made by the automated aid and to make significantly less omission and commission errors (failures to detect automation misses and false alarms, respectively; see also chapter 2.3.3.3) compared to the other experimental groups. In a driving simulator study by De Waard et al. (1999) participants perceived significantly higher risk when the driving task was automated and time headway to the car in front was fixed to a short interval (0.25 s) compared to a long interval (1 s). The higher level of perceived risk was paralleled by significant higher ratings of arousal and mental effort, presumably because the anticipated performance consequences of a system failure were more severe when time headway was short and drivers thus put more effort in system monitoring.

2.3.6.2 Summary

In summary, it appears that operators’ reliance on automated system is indeed mediated by their strategic decisions about the investment of mental effort. Especially highly reliable systems may induce the perception that the costs associated with intensive system monitoring are not outweighed by the expected benefits (e.g., early detection of system malfunctions). The above mentioned studies suggest that operators’ strategic decisions about mental effort investment and their reliance are very much influenced by the pay-off structure of a task. In multitasking conditions where system monitoring is only one of many concurrent tasks and where automation failures are not detrimental for overall system performance (as in many empirical studies) operators seem to manage effort in a way that optimises overall performance. Low automation reliability appears to encourage operators to invest more effort in system monitoring in order to yield an acceptable level of performance. Furthermore it can be suggested that operators try to adopt performance strategies that maximise the cost-benefit ratio between effort investment and performance output. This is indicated for example in the Parasuraman et al. (1993) and Bahner et al. (2008) studies in which those experimental groups that were supposed to have invested less effort in system monitoring did not show improved performance in concurrent tasks. Operators may tend to optimise performance rather than to maximise it. Thus, if performance is considered to be acceptable, operators may not try to invest additional effort in order to even improve performance, but rather to conserve processing resources in order to be able to deal with unexpected increases in task demands (see also a current literature review on the effects of automation reliability and task load on operators' performance, Wickens & Dixon, 2007). When task demands or performance costs increase - for example when automation failures occur or the consequences of such failures become more severe - operators will likely invest more effort in order to improve performance and to mitigate negative performance consequences.

2.4 Situation Awareness

2.4.1 What is situation awareness?

The term situation awareness has emerged from aviation psychology in the 1980s in order to account for human errors that occurred in connection with modern aircraft technologies and apparently resulted from pilots' insufficient understanding of the current situation, including a lack of awareness about the state of various automated system components. Situation Awareness (SA) thus is related to the ability of an operator to achieve and maintain a current and comprehensive understanding of the situation while interacting with an increasing number of automated systems. Meanwhile, SA has been applied to many areas in which operators act in complex dynamically changing environments and where automation of task processes becomes more and more prevalent (Durso & Sethumadhavan, 2008; Wickens, 2008). While the intuitive notion of "knowing what is going on around you" (Endsley, 2000b, p. 5) as a prerequisite for successful performance is most widely accepted by human factors researchers; the status of SA as a concept, its definition and measurement is still subject to an ongoing debate (e.g., Dekker & Hollnagel, 2004; Durso, Rawson, & Girotto, 2007; Flach, 1995; Gelau & Krems, 2009; Rousseau, Tremblay, & Breton, 2004; Sarter & Woods, 1991; Tenney & Pew, 2006).

The most influencing definition was given by Mica Endsley who defined SA as "the (1) perception of the elements in the environment within a volume of time and space, (2) the comprehension of their meaning and the (3) projection of their status in the near future" (Endsley, 1995b, p. 36). Other definitions of SA have mostly adhered to this conceptual understanding, although they differ in the proposed linkage of SA to other cognitive concepts (such as attention, perception, and working memory) and its integration in existing cognitive psychological theories (such as information-processing, decision-making). Maybe the most apparent divergence in conceptual definitions of SA refers to the *product* or *process* aspect of SA, i.e. whether SA is regarded as a state of knowledge – a snapshot of the operators' momentary mental representation of the situation (e.g., Endsley, 1995b), or as the cognitive processes necessary to achieve and maintain this mental representation (e.g., Smith & Hancock, 1995). The view of SA as a product or process has important implications for its operational definition and its measurement. Yet another view is taken by researchers who emphasise the interrelationship between product and process in postulating that the current mental representation affects the process of acquiring and interpreting new information in an on-going cycle (e.g., Adams, Tenney, & Pew, 1995, by referring to Neisser's perceptual cycle). This view of *SA as interrelated product and process* is hold in this work. SA can accordingly be defined in line with Sarter & Woods as "the accessibility of a comprehensive and coherent situation representation which is continuously being updated in accordance with the results of recurrent assessments" (Sarter & Woods, 1991, p. 52).

2.4.2 The role of top-down and bottom-up attentional processes in achieving and maintaining SA

A central role in this definition has the concept of a dynamic mental representation of the situation, which is also referred to as situation model following Van Dijk and Kintsch's (1983) notion in text comprehension (Baumann & Krems, 2007; Durso & Gronlund, 1999; Durso et al., 2007; Endsley, 1995b). This situation model is dynamic in the sense that it has to be continuously updated through the alternation of top-down and bottom-up attentional processes (Adams et al., 1995; Baumann & Krems, 2007; Durso et al., 2007; Endsley, 1995b, 2000b). The construction of the situational model is based on the perception of new elements in the environment which are then integrated into a coherent mental representation by means of pattern recognition (Durso & Gronlund, 1999; Durso et al., 2007). Pattern recognition happens through the activation of knowledge structures in long-term memory such as mental models and schemata whereby the current situation is connected to prior experiences of similar situations or configurations of elements (Baumann & Krems, 2007). This activation of stored knowledge in long-term memory serves as a basis for the comprehension of the current situation, as a mean for guiding subsequent sampling of the environment, and allows for anticipating of the progress of the current situation. Furthermore, the construction of the situation model is assumed to also activate related action schemata that may directly be implemented based on the recognition of typical situations (Klein, 1993). Baumann and Krems (2007) offer a detailed account of how situation awareness may be achieved and maintained while driving and how it may influence action selection by referring to Kintsch's Construction-Integration theory of text comprehension (Kintsch, 1998) and Norman and Shallice's theory of action selection (Norman & Shallice, 1986). McCarley, Wickens, Goh and Horrey (2002) provide a computational model of SA with particular emphasis on the role of top-down control of attention in achieving and maintaining SA.

The successful alternation of top-down and bottom-up control of attention is seen as a prerequisite for maintaining a high level of situation awareness (Adams et al., 1995; Durso et al., 2007; Endsley, 1995b, 2000b). The currency and comprehensiveness of the situation model is necessary for guiding attention to the relevant cues in the environment (top-down control of attention). Similarly, SA requires the sensitiveness for new information (particularly information that contradicts the momentary comprehension of the situation) in order to be able to react to changing situational demands. The successful interplay between top-down and bottom-up processes is especially important in highly dynamic tasks and environments such as driving. In driving, the momentary situation may change possibly in a few seconds depending on the behaviour of other traffic participants, demanding adaptive responses from the driver including the reformulation of action goals. Nilsson (2005) described very well the role of top-down and bottom-up processes while driving: First, “driving includes continuous monitoring of relevant parts of the environment. The driver has to know *where and when* to search (mainly look) for and acquire necessary information.” (p. 294) Second, “the driver must be able to distinguish necessary and important information from irrelevant information for the task (trip) at hand. Therefore, car driving includes *selection* of relevant information from a “noisy” whole, meaning that the driver has to know *what* information is important and of relevance in various occasions and situations.” (p. 294) Finally, because driving is a highly dynamic task and depends on the interaction with other road traffic participants, there exist various degrees of uncertainty. Therefore, “car driving includes readiness for unexpected situations. The driver has to be prepared to revise and change planned actions, even if they are based on ‘correct’ predictions and interpretations based on previous knowledge and experience, because situations not always develop accordingly and ‘as usual’.” (p. 294) Whereas the first two points made by Nilsson highlight the role of expectancies or top-down processes in directing attention, the third point accounts for the importance of bottom-up processes in updating the situation model.

2.4.3 Automation and situation awareness

Automation results in changing task demands and therefore affects the human control processes that are normally involved when operators perform a task manually. Because of the goal-driven nature of attention, automation can also be expected to influence the control of attention as a central mechanism that enables humans to perform tasks. Automation may have positive as well as negative effects on operators’ SA.

2.4.3.1 The enhancement in SA due to lower mental workload with automation

The positive effects of automation on SA are generally attributed to its potential to relieve the operator from mental workload. The process of building up and maintaining a high and accurate level of SA can be considered a resource-intensive process (Endsley, 1995b; Wickens, 2001). Working memory plays a central role in this regard. As Baumann and Krems (2007) noted, “working memory resources are necessary for associating perceived elements in the environment with knowledge stored in long-term memory, for integrating these new elements in the current situation model, for removing irrelevant elements from the situation model, for keeping the information in the situation model available for the selection of appropriate actions, for monitoring the selection and execution of actions and so on” (p. 259). The limited capacity of working memory therefore acts a bottleneck in achieving and maintaining a high level of SA. Thus, automation may lead to higher levels of SA by reducing mental workload and enabling the operator to invest more cognitive resources in the maintenance of SA. Empirical evidence for the beneficial effect of automation on SA was for example found in a study by Ma and Kaber (2005) who studied the effects of driving with an ACC system and concurrent cell-phone use on drivers' mental workload and SA. A product-oriented, memory-based measure of SA (SAGAT - Situation Awareness Global Assessment Technique; Endsley, 1995a) was used in order to assess participants' responses to queries referring to Endsley's three levels of SA after freezing the simulation. Level 1 *(perception)* queries required recalling car locations and colours of traffic signs they had passed. Level 2 *(comprehension)* queries required participants to identify necessary driving behaviours in order to improve their following behaviour in the current situation (e.g., acceleration, braking etc.). Level 3 *(anticipation)* queries required participants to project times to certain events (e.g., the time to the next turn etc.). Ma and Kaber (2005) found that driving with ACC reduced drivers' mental workload compared to driving without ACC and that SA accordingly increased across all three levels. This result was interpreted in the way that the ACC decreased demands associated with continuous speed and headway control and therefore allowed drivers to invest more resources in the maintenance of SA.

Support for the view that reductions in mental workload help operators to maintain higher levels of SA was also found in a study by Endsley and Kaber (1999). Participants reported lower workload when the decision-making aspect of a simulated radar-monitoring task was automated, and their level 2 SA was superior compared to other forms of human-automation control allocations. However, this higher level of SA did not automatically lead to better performance when control was completely returned to participants at different times during the experiment.

Other research has shown however that reduction of mental workload does not necessarily lead to improved SA. Kaber and Endsley (2004) found that improvements in SA with intermediate levels of automation in a dynamic control task were not related to reductions in mental workload. Endsley and Kiris (1995) also found that changes in level 2 SA (situation comprehension) with varying levels of automation in a decision task were unrelated to participants' reported mental workload.

These findings altogether suggest that other factors related to cognitive processes and to operators' involvement in the task play a role in achieving and maintaining SA.

2.4.3.2 Vigilance-related impairments in SA and motivational factors

The reduction of mental workload is not necessarily beneficial for human performance, as was already described in chapter 2.3.5.2. Especially high automation levels and highly reliable systems minimise the demands for manual control and may therefore lead to underload, a decline in arousal and impair the operator's ability for vigilant monitoring. The maintenance of SA, i.e. the need to continuously update the situation model however requires sustained attention. Thus, in the same way as automation may lead to impairments in sustained attention, it may also impair SA.

According to Endsley (1996), overreliance on automation and/or complacency is the primary cause for vigilance-related impairments in SA. Thus, she assumes that complacency and overreliance reduce the operator's efforts for vigilant monitoring and for continuously updating the situation model. In other words, according to Endsley, the primary reason for vigilance-related impairments in SA is motivational in nature and results from the decision of the operator to invest fewer cognitive resources in maintaining SA than would be optimal. Although it can be assumed that both reliance on automation and complacency may negatively affect SA, the nature of the processes leading to impairments of SA may be different. This argues for keeping the two concepts and its hypothesised causal relation to SA distinct from another. Complacency is an attitude towards automation and may result in a high level of reliance, i.e. in the allocation of a considerable amount of control to an automated system. Whereas the impairment of SA due to a complacent attitude can be considered to be clearly motivational in nature, changes in SA due to reliance can be assumed to be caused both by cognitive and motivational factors. Cognitive factors can be referred to as changes in information processing and attentional control related to the allocation of control to a system. Motivational factors are related to operators' attitudes towards automation that influence their intention to rely on a system (such as trust and complacency).

The impairment of SA due to complacency appears to be a widely acknowledged fact in the literature, although it has been rarely empirically tested. Lower levels of SA are often attributed post-hoc to a high level of trust in a system and as a consequence, complacency. Normally however, neither trust or complacency (e.g., by assessing operators' system monitoring strategies; cp. Moray & Inagaki, 2000) are tried to be measured directly. Thus, there is little empirical evidence supporting the assumed negative relationship between complacency and SA, although it has some plausibility.

2.4.3.3 The impairment in SA due to changes in cognitive processes

Automation and reliance (as the action of allocating control to a system) can be assumed to influence the operator's cognitive processes underlying performance. Automation changes the operator's task(s) usually from active controlling to monitoring of the automated functions. Consequently, automation changes task performance requirements and accordingly, affects human information-processing and the control of attention as a basic mechanism for resource allocation. It was already mentioned above that the successful alternation of top-down and bottom-up control of attention is a prerequisite for the maintenance of a current and comprehensive situation model.

Changes in information-processing and in the control of attention may result from changes in the task goal structure and/or the disruption of control processes under automated conditions.

Changes in the task goal structure due to automation

Automated systems always rely on some sensory input and internal information-processing algorithms that are applied to this sensory input. That is, some sort of information analysis is an integral part of all automated systems regardless of which function they are designed to automate. Automation therefore changes or reduces demands associated with the acquisition and analysis of certain types of information for the operator as these functions are (more or less sufficiently) performed by the system. It can be expected that through these changes in information-processing demands the informative value of certain perceptual information (cp. McCarley et al., 2002) changes in that it may become less relevant for performance under automated conditions. For example, when driving with a lane keeping assistance system, perceptual information related to lateral control may become less relevant over time. As a consequence, fewer attentional resources may be allocated to the processing of this information which may in turn hinder appropriate bottom-up control of attention, for example, when unexpected situations occur that demand for the driver's intervention (e.g., obstacles on the road that are not recognised by the lane keeping assistance system).

Hoc et al. (2006) performed a driving study in which they investigated whether driving with a Heading Control (HC) system affected drivers' processing of visual information necessary for either lateral or longitudinal control of the vehicle in curves. Specifically, they compared drivers' perception of visual information located either near to the tangent point of a curve (assumed to be critical for lateral control of the vehicle) or located in the straight-ahead visual field (assumed to be critical for speed adjustment). For this purpose, they placed non-familiar advertising logos near to these two interesting points. After having passed the curves, they asked drivers to stop and to give subjective ratings of how sure they were about having seen these logos. Results showed that drivers had higher recognition scores when driving with the HC system than without, which was however mainly attributed to technical problems and an effect of the procedure (Apparently, certainty judgements were only available for six out of 12 drivers who drove first without assistance and then with the HC system. Thus it was assumed that these drivers were more prepared to recall the logos when they encountered the situation a second time while driving with the HC system.). Interestingly (although no inferential statistics are reported), when driving with the HC system, participants judged to have seen the logos placed straight ahead with higher certainty than the logos located near to the tangent point. The reverse was found for the manual driving condition, for which certainty scores were higher for logos located near to the tangent point than for logos located straight ahead. Although doubts may be raised regarding the subjective nature of the measurement, results seem to be in accordance with the assumption that active processing of information related to an automated subtask (in this case: lateral control) may be reduced because it is no longer relevant for performance of the remaining task components (in this case: longitudinal control). It must be added that Hoc et al. (2006) attributed this effect not to changes in cognitive processes, but to complacency.

The disruption of control processes due to automation

On the other hand, automation intervenes in the control processes that are normally involved when operators perform a task manually and therefore interrupts the feed-forward and feed-backward control mechanism that are necessary for keeping an operator actively involved in a task. Recently, the driving task has been described as a set of simultaneous, interrelated and layered control processes (Engström &

Hollnagel, 2007). Within the Extended Control Model (ECOM) four simultaneously active layers of control are distinguished: tracking, regulating, monitoring, and targeting. The outputs of a higher control loop form the objectives (goals, criteria) of a lower control loop. For example, activities initiated at the regulating level (such as lane changes) affect the activities at the tracking control loop (e.g., control of speed and headway). In turn, the input of the lower levels is required for efficient control on the higher levels. Today's driver assistance systems mostly automate activities at the tracking control loop, such as longitudinal control (ACC) or lateral control (HC). However, the automation of tracking activities is likely to make control on the higher levels more difficult because important information is filtered out from the dynamic control process (Hollnagel & Bye, 2000; Hollnagel & Woods, 2005). As Hollnagel points out, the "efficiency of regulating requires the input from the tracking activities. If these therefore are heavily automated, regulating is likely to suffer, even though the tracking activities themselves may be efficiently carried out by automation." (Hollnagel, 2002, p. 20) The lacking feedback from the tracking control loop may result in failures in bottom-up control of attention, such that information demanding for changing objectives at the higher control levels is not sufficiently processed. Referring to SA and Norman and Shallice's (1986) theory of action selection, the relevant information may not become activated and therefore not available in the situation model and thus, is not able to trigger the appropriate action schemata.

Several studies were carried out in order to investigate the effects of different levels of automation on operators' SA. Those studies have yielded in part conflicting results, demonstrating that there is no simple relationship between the degree of task automation and SA, mental workload, and task performance. Results in fact appear highly dependent on the concrete realisation of different levels of automation in specific task settings, the experimental procedure (e.g., duration of automated vs. manual control periods), the performance requirements (e.g., reward structure), and the methods used to assess SA and mental workload for example.

Endsley and Kiris (1995) investigated the effect of increasing levels of automation in a decision-making task on performance, SA and mental workload. They found that increasing automation of decision-making was related to increasing impairments in level 2 SA (situation comprehension), but not in level 1 SA (situation perception). Although participants seemed to have perceived the relevant information, they did not develop a higher-level understanding of the situation as the automation increasingly relieved them from the evaluation of decision alternatives and from action implementation. The impairment in level 2 SA was mirrored by the time participants needed to reclaim manual control of the task when the automated decision aid failed; i.e. longer task times after an automation failure were associated with lower levels of SA. Conflicting results were found by Kaber and Endsley (2004) who compared six levels of automation (from full manual control to full automation) with different degrees of computer assistance in four basic task functions (monitoring, generating, selecting, implementing) involved in a simulated radar-monitoring task. They found that computer assistance in monitoring (sensory processing) and action implementation functions of the task resulted in worse (level 2) SA as compared to assistance in decision-making functions and full automation of the task. However, in manual control periods (described as automation failures), performance was better when assistance was previously applied to sensory functions and action implementation (those levels associated with lower SA). Endsley and Kaber (1999) found that computer assistance primarily in action selection and implementation yielded higher-level 2 SA, but mostly worse level 3 SA and were associated with relatively long recovery times in manual control periods.

Kaber, Perry, Segall, McClernon, and Prinzel (2006) investigated the effects of automation of different functions (information acquisition, information analysis, decision-making, action implementation) on SA, mental workload, and performance in an air traffic control simulation. Automation of information acquisition resulted in highest level 1 SA (perception) and automation of action implementation resulted in lowest level 1 SA. Performance was best when automation was applied to information acquisition and action implementation; however, during intermediate manual control periods, performance was worse when automation was formerly applied to decision-making and action implementation.

Summary

In summary, research on the impact of automation on operators' SA has not yielded consistent results. It is predicted that automation affects human information processing and disrupts the top-down and bottom-up control of attention. The active processing of information and the successful interplay of top-down and bottom-up attentional processes are however seen as a prerequisite for maintaining an accurate and comprehensive understanding of the current situation. However, research indicates that the effects of automation on cognitive processes are complex and that automation may not necessarily reduce, but also enhance SA. Several reasons may account for this finding. First, 'automation' is a broad term and means different things in different task contexts. Performance requirements of particular types and levels of automation may be different across task contexts and even within one particular context, automation may be realised in different ways (e.g., through different designs of the human-machine interface) associated with different demands for human information processing. Second, knowledge-based and process-oriented measures may be sensitive to qualitatively different aspects of SA or may dissociate under specific conditions. Most research has focused on changes in operators' state of knowledge rather than on the cognitive processes necessary to achieve this state of knowledge. Often it was found that particular types of automation led to improved SA under normal conditions, but led to degraded operator performance (e.g., longer recovery times) during intermediate manual control periods. One explanation for this intuitively contradictory finding could be that the cognitive processes involved in the maintenance of SA were disrupted by automation, but in the same vein did automation help operators to maintain an appropriate state of knowledge about the current situation by supporting them in information analysis or providing them with decision support for example. However, taking away this additional support may have led to impaired performance in periods of manual control.

2.5 Aim of Research and Research Hypotheses

2.5.1 Aims of this study

The aim of this PhD work is to comprehensively study changes in the level of drivers' active engagement in the driving task as a function of the increasing automation of the lateral control subtask. It is predicted that drivers' active engagement in the driving task declines the higher the extent to which the human control processes involved in lateral control of the vehicle become automated. Changes in drivers' level of engagement in the driving task are assessed by referring to two established psychological concepts: drivers' reliance on a lane keeping assistance system and drivers' situation awareness.

Reliance was defined in chapter 2.1 as the deliberate act of allocation varying levels of control to an automated system. Increasing reliance on an automated system is associated with less active task engagement attributable to operators' increased preparedness to hand over responsibility for maintaining task performance to the system, and to invest less effort in critically monitoring and examining the appropriateness of the system's actions or behaviour.

Situation Awareness (SA) refers to the maintenance of a current and comprehensive mental representation of the actual situation (Endsley, 1995b, 2000b). It is assumed that SA increasingly suffers the more operators rely on an automated system, i.e. the more control they allocate to an automated system (Endsley, 1996). Thus, SA is expected to deteriorate the less operators are actively engaged in a task due to changes in cognitive, motivational and energetic processes associated with the allocation of increasing levels of control to an automated system (feed-back loop of reliance, see Figure 1 on page 37). Difficulties in the process of maintaining SA are however also expected to result from the automation of human control processes 'per se' due to changes in human information processing and in the control of attention, as well as due to changes in energetic and motivational processes.

The ultimate goal of this work is to study drivers' reliance on lane keeping assistance systems and drivers' SA in relation with their attitudes and changes in their energetic and motivational processes as a response to the increasing automation of the lateral control driving subtask. Based on an extensive review of the research literature a theoretical framework was proposed that relates changes in operators' reliance and SA to changes in underlying human cognitive, energetic, and motivational processes. Previous research has not sufficiently taken into account these mediating influences of human adaptation processes in response to automation when studying operators' situation awareness and reliance on automation. In most studies, if any, only single mediating variables were considered explicitly. Consequently, those studies generally do not provide sufficient explanation for the *nature* of the observed changes in operators' level of active task engagement. Furthermore, the same dependent variables have been used as an equivalent measure for a range of hypothesised underlying factors contributing to operators' out-of-the-loop performance under automation. For example, operators' impaired detection performance for automation failures was equivalently used as an indicator for their overreliance on automation, their excessive trust in automation capabilities, their complacency, their loss of SA, their reduced vigilance and so on. The aim of this study is to assess the differential impact of cognitive, energetic, and motivational processes on changes in drivers' level of active task engagement in response to the increasing automation of the lateral control task *by monitoring a range of objective and subjective dependent measures*.

Furthermore, two different process-oriented performance-based measures will be used in order to study the effects of the increasing automation of the lateral control task on drivers' SA and their reliance on lane keeping assistance systems. Although researchers generally assume a negative relationship between operators' reliance on automation and their SA, no attempts were made up to now to study these two concepts in conjoint. Apart from that, the present theoretical framework assumes that increasing automation 'per se' may negatively affect operators' SA (independently from operators' reliance on it) primarily due to changes in cognitive processes, but also due to changes in energetic and motivational processes in relation with a system's increasing intervention in human control processes.

Up to now, empirical research on reliance and on SA has followed rather distinct approaches. While research on reliance on automation has almost exclusively focused on the effect of varying levels of automation reliability on reliance, research on SA has focused on the effect of different types and levels of automation on SA. Research on both concepts is however driven by researchers' interest in the *changes in operators' active task engagement as a function of how completely they are taken out of the control loop by automation*. Both highly reliable automation and high-level automation minimise the need for operators to (re)engage in manual control and therefore the need, to stay actively involved in a task. A review of empirical studies on operators' reliance on automation revealed that the reduced necessity for manual intervention with highly reliable automation affected human energetic and motivational processes, as well as operators' attitudes towards automation. In general, operators were found to increasingly rely on an automated system the higher its reliability. Furthermore, the more operators relied on an automated system, the less effort did they invest in monitoring the automated system's action and in verifying its behaviour (e.g., Bagheri & Jamieson, 2004; Bahner et al., 2008; Muir & Moray, 1996). The major assumption in this work is that the level of automation has a similar effect on operators' reliance as automation reliability by affecting underlying motivational and energetic processes in a similar way. Thus, it is assumed that drivers increasingly rely on a lane keeping assistance system the higher the degree of assistance and the lower the manual control demands.

Of special interest in this study is the *change in drivers' reliance as a response to the experience of functional system limits*. As already mentioned, operators' reliance on automation has been studied up to now only as a function of varying levels of automation reliability. It was found that in general, operators' trust in and their reliance on automation declined as a response to the occurrence of automation faults or failures (e.g., Lee & See, 2004; Muir & Moray, 1996). In this study, the lane keeping assistance systems were completely reliable within their functional scope throughout the entire study. However, drivers encountered particular driving situations during the second half of the experiment that required them to react with respect to lateral control of the vehicle (e.g., by drawing aside) in order to avoid collisions with other traffic. These situations lay outside the functional scope of the lane keeping assistance systems as they were not designed to detect those situations. Consequently, the assistance provided by the systems (i.e., supporting drivers to stay within the driving lane) was not adequate in these situations. It is supposed that drivers' experiences of functional system limits has a similar negative effect on their trust in and their reliance on lane keeping assistance systems as operators' experiences of system faults in previous studies. Furthermore, it is assumed that the effect of functional system limits on trust and reliance are more pronounced for higher levels of lane keeping assistance, as predicted by theoretical frameworks (Lee & See, 2004; Parasuraman et al., 2000; Parasuraman & Wickens, 2008). This is because the inadequacy of a system's behaviour (as in the case of functional system limits) is associated with more adverse consequences for performance with automation exerting greater control over action regulation functions compared to automation exerting less control over human control processes. The higher adversity of performance consequences associated with functional limits of higher-level automation is supposed to be due on the one hand to the greater magnitude of these consequences, and on the other hand to the higher degree of their inevitability. With low-level automation operators have the possibility to ignore the additional information or the warnings provided by this automation, whereas even a low level of reliance on higher-level automation is associated with the allocation of a substantial amount of control to automation.

In line with recent theoretical frameworks (Lee & See, 2004; Meyer, 2004) it is assumed that operators' reliance on automation is not a static behaviour, but that operators adapt their reliance dynamically to the perceived changes in the automation's performance, in task demands as well as to their energetic state. As suggested by empirical evidence, it is predicted that changes in operators' reliance are mediated by changes in attitudes towards automation (e.g., trust) as well as by changes in motivational and energetic processes. Thus, it is assumed that the greater decline in drivers' reliance on higher levels of lane keeping assistance as a response to functional system limits is paralleled by changes in drivers' trust and in motivational and energetic processes. Mental effort regulation is assumed to play a dominant role in mediating drivers' reliance. Specifically, it is assumed that drivers invest more effort in monitoring the lane keeping assistance systems' actions in order to be able to prevent large deteriorations in driving performance and safety associated with the inadequacy of the systems' behaviour in those critical situations. Furthermore, drivers are expected to invest more effort in maintaining an acceptable driving performance when faced with critical situations (partly representing functional system limits). The increase in driving task-related effort and in monitoring the lane keeping assistance systems' actions should be more pronounced for drivers driving with a high level of lane keeping assistance. Thus, the experience of functional system limits should lead to drivers′ more active engagement in the driving task (especially when the level of lane keeping assistance is high) indicated by their decline in reliance on the lane keeping assistance systems and their increase in driving task-related effort.

2.5.2 Methodological approach

An advanced driving simulator was used to investigate changes in drivers′ level of active engagement in the driving task as a function of the increasing automation of the lateral control driving subtask and their experience of functional automation limits. A driving simulator was chosen in order to minimise safety risks in relation with drivers′ experience of functional limits of the lane keeping assistance systems. A second major reason for using a driving simulator instead carrying out a field study in real traffic was the larger experimental control over procedures and over the manipulation of independent variables which was regarded as essential in this study. Furthermore it was supposed that the eventually lower safety risks perceived by drivers in the simulator compared to real traffic conditions would increase the chance of observing an effect of the level of automation and of drivers′ experience of functional system limits on their reliance on the lane keeping assistance systems and on their SA.

2.5.2.1 Levels of automation of the lateral control driving subtask

Changes in drivers′ active task engagement were investigated as a function of the increasing automation of the lateral control subtask. Automation of lateral control was chosen as a prototypical example for the increasing automation of the driving task because it allowed for comparably "tiny" manipulations of the level of automation with experimental control over possible side effects arising from the design of the human-machine interface for different levels of automation. Thus, the different levels of automation were designed to be comparable in terms of the visual and cognitive demands they would impose on the driver.

Three different levels of automation of lateral control were implemented in this driving simulator study. A *low* and a *high level of automation* were realised by two lane keeping assistance systems that were designed to prevent drivers from unintentional departures from the driving lane. The two lane keeping assistance systems employed were comparable to systems that are currently available or introduced to the market in terms of their human-machine interface design. The two levels of automation realised by these two lane keeping assistance systems will be referred to hereafter as a high and a low level of lane keeping assistance. A *high level of lane keeping assistance was realised by a Heading Control (HC) system* that actively intervened in drivers′ lateral control of the vehicle by applying counter forces on the steering wheel in order to guide the vehicle within the centre of the lane. A *low level of lane keeping assistance was realised by a Lane Departure Warning (LDW) system* that warned drivers by a haptic feedback on the steering wheel (vibration) as soon as they were approaching the left or right lane marking. Thus, the LDW system basically did not relieve drivers from lateral control, but informed them as soon as task-related safety limits were reached (near-crossing of the lane boundaries) indicating the need for them to take corrective steering actions in order to stay within the driving lane. A *control group of drivers drove without any lane keeping assistance* (corresponding to a zero level of automation, referred to hereafter as *No Assistance*).

According to Parasuraman et al.'s (2000) model of different types and levels of automation, the LDW system corresponded to a high level of automation of information analysis (with no automation of decision selection and action implementation), whereas the HC system corresponded to a medium level of automation of action implementation (see Figure 2).

Although it is possible to locate the different levels of lane keeping assistance used in this study on the dimensions of the Parasuraman et al. (2000) model, it appears that automation of the lateral control driving subtask is not necessarily comparable to automation of other tasks, even if the same (types of) human information processing functions are automated to the same extent. The tasks usually used in automation research are often characterised by much higher cognitive demands than the driving task in terms of resource-intensive conscious processing of information, for example when operators have to operate a complex process control simulation and are responsible for fault diagnosis and management (diagnostic reasoning) in different subsystems.

In comparison to many other task used to study the effects of automation on human performance, the driving task is characterised by a much higher temporal and spatial dynamic of task elements and task constraints (in terms of traffic regulations) that a drivers has to take into account while pursuiting a number of concurrent goals on different control layers. To a large part, driving requires the continuous extraction and extrapolation of visual information and its interpretation (more or less conscious) in terms of these goals and the actual affordances of the driving situation. Drivers' difficulties associated with the supervision of driving task automation arise not that much from the reduced capability to observe what the automation does (because of a physical separation from the controlled process), but rather from the requirement to keep continuously track of the rapid and dynamic (often subtle) changes in environmental conditions in order to be able to immediately take-over manual control when the automation does not behave as it should. Thus, the need for drivers' active engagement in the driving task particularly arises from the dynamic nature of the driving task and from the need for constant vigilant monitoring, processing and evaluation of environmental information in order to reclaim control when the automation fails.

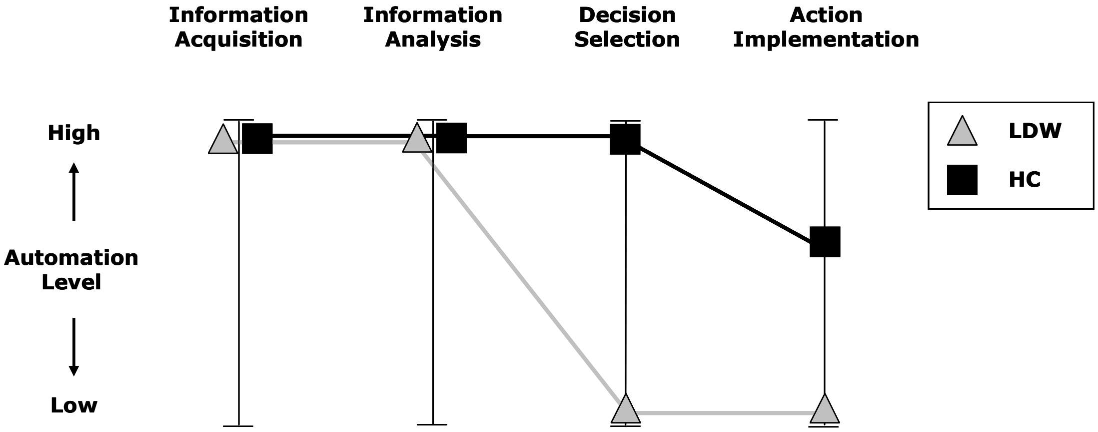


Figure 2. Level of automation of four human information processing functions realised by the Heading Control (HC) and Lane Departure Warning (LDW) system according to Parasuraman et al.'s (2000) model of types and levels of automation.

2.5.2.2 Measurement of drivers' reliance on the lane keeping assistance systems

A process-oriented performance-based measure was used in order to study drivers' reliance on a lane keeping assistance system. Empirical evidence suggests that reliance on automation is not a stable behaviour in response to specific properties of the automation, but that reliance varies dynamically as a function of the actual performance of an automated system, the momentary environmental context (situational demands, costs and benefits associated with joint human-system performance consequences), and individual predispositions of the operator that combine to influence energetic and motivational processes. Performance-based measures of reliance must therefore be sensitive to operators' dynamic allocation of control to automation as a function of changes in context and in the automation's performance. In previous studies, the proportion of time operators chose to engage automation vs. preferred to perform a certain task manually, or their automation monitoring behaviour (portion of information sampled, time used to verify automation decisions) were used as process-oriented performance-based measures of reliance. In this study, drivers' reliance on a lane keeping assistance system is measured by drivers' preparedness to allocate their visual attention away from the road scene to the visual display of a secondary task mounted in the centre console of the vehicle. Drivers' increased preparedness to allocate their visual attention away from the forward road scene (corresponding to the most relevant source of information while driving) is thus viewed as an indicator for their allocation of a higher level of control to the lane keeping assistance system and their less active engagement in the driving task.

The visually demanding secondary task served two main purposes. First, it should provoke a greater variation in drivers' lateral control of the vehicle (e.g., Engström, Johansson, & Östlund, 2005; K. L. Young, Regan, & Lee, 2008; Zhang, Smith, & Witt, 2006; Zwahlen, Adams, & DeBald, 1988) in order to enable them to experience the actions and behaviour of the lane keeping assistance systems. The experience of a system's behaviour is a prerequisite for operators to develop a certain level of trust in its performance which in turn is assumed to influence their reliance on the system (Lee & See, 2004). Second, the secondary task was designed to foster drivers' reliance on the lane keeping assistance systems by generating a resource-demanding dual-task situation (cp. Rudin-Brown & Noy, 2002). The aim was to manipulate task demands in a way that encourages drivers to rely on the lane keeping assistance systems as a strategy to limit the expenditure of mental effort in the driving task while maintaining or even improving concurrent task performance. Theoretical frameworks assume that operators' reliance on automation is based on a utility judgement regarding the costs and benefits associated with the allocation of different levels of control to an automated system (Dzindolet et al., 2002). The major benefit associated with higher levels of reliance appears to be the potential to conserve mental effort for operators while maintaining or even improving performance due to the automation support. Accordingly, empirical studies yielded evidence that operators' reliance on automation was more pronounced when task demands were high compared to low (McFadden, Giesbrecht, & Gula, 1998; Parasuraman et al., 1993; Wickens & Dixon, 2007). Thus, the secondary task should promote changes in drivers' reliance to occur during a limited time of exposure to the lane keeping assistance systems, while at the same time drivers' strategies to allocate their visual attention between the road scene and the visual display of the secondary task served as a measure for their reliance on the lane keeping assistance systems.

Studies on driver distraction indicate that drivers' time-sharing strategies for allocating their visual attention between the road scene and the visual display of an in-vehicle task are influenced by the visual complexity of the information presentation and the driving task demands (Lansdown, 2001; Tsimhoni & Green, 2001; Victor, Harbluk, & Engström, 2005; Wierwille, 1993). If costs for visual reorientation are high, as for example when in-vehicle tasks place high demands on visual search (e.g., reading a rotating map of a route navigation display), the duration of drivers' single glances to the in-vehicle display increases (Hoffman, Lee, McGehee, Macias, & Gellatly, 2005; Lansdown, 2000; Tsimhoni & Green, 2003; Victor et al., 2005). On the other hand, drivers usually react to increasing driving task demands with a higher frequency of shorter glances to the in-vehicle display (Lansdown, 2001; Tsimhoni & Green, 2001, 2003; Wierwille, 1993; Wikman, Nieminen, & Summala, 1998).

The secondary task used in this study was a visual search task of medium complexity developed in the European HASTE project (Carsten et al., 2005; Engström et al., 2005). It was expected that drivers tend to respond to this secondary task if possible with single long glances to the visual display because of the cost of increased response times associated with an intermittent glance strategy (i.e., with subsequent alternating glances to the road scene and to the visual display). It was further expected that drivers' increasing reliance on the lane keeping assistance systems should be reflected in a higher extent of their adoption of that strategy. Thus, drivers' increasing allocation of control to a lane keeping assistance system should be reflected by longer eyes-off-the-road times and a higher proportion of glance time to the in-vehicle display during secondary task performance compared to lower levels of reliance. Furthermore it was hypothesised that drivers' reliance on the lane keeping assistance systems would increase with higher levels of assistance because the increasing lane keeping support should reduce driving task demands and their influence on drivers' visual attention allocation strategies.

This reduction of driving task demands was assumed to result partially from the haptic feedback of the lane keeping assistance systems that informed drivers about the vehicle's position in the lane even when they directed their visual attention away from the road scene. The LDW system provided this feedback only near to the lane boundaries in form of a non-directional binary warning (on/off) indicating the need for a corrective steering action of the driver in order to prevent an unintentional lane departure. The HC system provided continuous haptic feedback in the form of counter forces on the steering wheel that varied in strength according to the magnitude of the vehicle's lateral displacement from the centre of the lane. Previous studies provided evidence for the reduction of visual demands associated with increasing levels of haptic feedback provided by lane keeping assistance systems by using the occlusion paradigm (Buld & Krüger, 2002; Griffiths & Gillespie, 2005).

2.5.2.3 Measurement of drivers' situation awareness when driving with varying levels of lane keeping assistance

Following Sarter and Wood's (1991) definition, SA is viewed as interrelated product and process in this work (see chapter 2.4.1). Thus, it is assumed that the current mental representation of the situation affects the process of acquiring and interpreting new information in an on-going cycle. Automation is generally expected to negatively affect operators' SA (e.g., Endsley, 1996; Endsley & Kiris, 1995; Sarter & Woods, 1992). However, the mechanisms by which automation impairs the cognitive processes necessary to achieve and maintain SA are currently not well specified. It is predicted here that automation interrupts the successful alternation of top-down and bottom-up attentional processes which is necessary to build up and maintain a comprehensive and current mental representation of the situation. Specifically, it is assumed that the informative value of certain perceptual information changes because this information becomes less relevant for performance under automated conditions. It is expected that with increasing automation of the lateral control subtask drivers allocate fewer attentional resources to the processing of visual cues relevant for lateral control of the vehicle. It is further hypothesised that the reduced processing of information relevant for lateral control becomes apparent in failures in the bottom-up control of attention in situations in which drivers cannot rely on the lane keeping assistance systems' actions because they are providing inadequate support. Thus, it is assumed that failures in the bottom-up control of attention result in an insufficient processing of visual cues demanding for drivers' intervention on the lateral control level, leading to drivers' delayed reaction in situations representing functional system limits.

This study represents one of the first attempts of studying content-specific changes in drivers' SA by employing process-oriented performance-based measures of SA. The SA measures used in this study can be described as *implicit* performance measures using response latency and response magnitude as the major dependent variables (Durso & Gronlund, 1999; Durso et al., 2007). Although decision-making and response selection processes are not viewed as integral part of SA (though supporting it), it is assumed that the initial point of drivers' reaction can be used to draw conclusions about automation-induced changes in the cognitive processes which are involved in the incorporation of specific environmental information in the actual mental representation of the situation. As Durso et al. (2007) point out, "methods that use time as a dependent variable are a step toward investigating the cognitive processes that underlie situation comprehension" (p. 171). A similar approach was recently used by Baumann, Petzoldt, Groenewoud, Hogema, and Krems (2008) who measured the effects of different types of cognitive loading tasks on drivers' response times to critical driving situations that required different degrees of anticipation. Also, the current approach to measure SA resembles the event-detection paradigm used in previous studies on driver distraction (e.g., Strayer, Drews, & Johnston, 2003).

In previous studies, mostly knowledge-based measures were used in order to assess changes in drivers' SA when driving with different driver assistance systems (Ma & Kaber, 2005; Ma & Kaber, 2007; Stanton & Young, 2005). However, knowledge-based measures of SA such as SAGAT (Endsley, 1995a, 2000a) are criticised for relying too heavily on memory and for assessing operators' recall performance after an interruption rather than their SA (Durso & Gronlund, 1999). A major reason for not applying knowledge-based measures in this study was that they were considered as too intrusive because of their potential to shift drivers' attentional focus after drivers would have recognised the characteristics of the probes (Rauch, Gradenegger, & Krüger, 2006). Another important reason for not using SAGAT-like techniques lay in the conceptualisation of SA as interrelated product and process. It was assumed that content-specific changes in drivers' SA would be reflected in differences in the timeliness (and eventually also in the appropriateness) of drivers' reactions to driving situations that required bottom-up controlled processing of information differing in their relevance for the performance of the remaining non-automated portions of the driving task. Drivers' delayed and more abrupt reactions to situations that require processing of information either relevant for lateral or longitudinal control of the vehicle were assumed to be indicative for their lower SA with regard to that respective type of information. Thus, content-specific changes in drivers' SA when driving with different levels of lane keeping assistance were assessed by the relative point in time and the abruptness of drivers' reactions to driving situations that differed in their processing demands for information either related or unrelated to lateral control.

For this purpose, eight driving situations were designed that occurred unexpectedly for drivers during the second half of the experiment. These situations should elicit testable responses. *Testable responses* are responses to "isolated, experimentally controlled events that can not be anticipated through any means other than good situation awareness, and that require a discernible, identifiable action (or set of actions) from the operator" (Pritchett & Hansman, 2000, p. 201).

The eight situations corresponded to different categories (see Table 1). Seven situations were safety-critical in that they required drivers to react in order to avoid a collision. The eighth situation assessed drivers' SA by their tendency to follow the erroneous behaviour of a vehicle driving in front of the drivers ("car-following"). The seven safety-critical situations required drivers to react primarily by adjustments in lateral or by adjustments in longitudinal control in order to avoid a collision. This means that drivers could principally react to these situations with respect to both lateral and longitudinal control; however, the situations were designed in a way that demanded necessary adjustments in either one of both driving subtasks in order to prevent a collision. Furthermore, the situations differed with respect to the time horizon available for drivers’ perception and processing of the relevant visual cues and for subsequent response planning and execution (small or large). Also, the situations placed different demands on higher cognitive processes by requiring either a low or a high amount of anticipation of the behaviour of other traffic participants and of the future development of the situation.

Table 1. Categories of critical driving situations used in this study in order investigate content-specific changes in drivers' situation awareness when driving with different levels of lane keeping assistance (Numbers in brackets refer to the number of critical situations occurring in this category).

Safety-critical (7)		
	Time horizon available	
Anticipation demands	small	large
high		“anticipatory” lateral (1)[a] longitudinal (2)
low	“unforeseeable” lateral (1)[a] longitudinal (1)	“foreseeable” lateral (1) longitudinal (1)
Not safety-critical (1)		
“car-following” longitudinal (1)		

[a] Situation categories in which an effect of the level of lane keeping assistance was expected.

These different categories of situations were designed in order to investigate hypothesised *changes in the depth of processing* for visual information either relevant for the subtask under varying levels of automated control (lateral control) or for the subtask still under manual control (longitudinal control). It was predicted that the increasing automation of the lateral control task leads to a reduced processing of visual information relevant for the lateral control task because of the disruption of bottom-up attentional processes. This in turn should become apparent in drivers’ delayed and more abrupt reactions to situations that demanded their intervention primarily on the lateral control level. Delayed and more abrupt driver reactions were expected in “unforeseen” situations with a relatively small time-horizon available for information processing and response selection, and in situations that required a higher amount of anticipation (deeper processing). No differences in reaction times and in the abruptness of drivers' reactions were expected for situations that required drivers to react with respect to longitudinal control, and for situations that *did* require drivers to react on the lateral control level while being unambiguous ("foreseeable") and leaving drivers enough time to perceive and interpret visual cues and to plan ahead their reaction.

2.5.3 Research hypotheses

2.5.3.1 Independent variables

Two main independent variables were manipulated in this driving simulator study: the level of lane keeping assistance and the occurrence of critical driving situations that partly represented functional system limits (see Table 2).

The *level of lane keeping assistance* was varied between-subjects. One group of drivers drove with a *Heading Control (HC)* system, representing a high level of lane keeping assistance. A second group of drivers drove with a *Lane Departure Warning (LDW)* system, representing a low level of lane keeping assistance. A third group of drivers drove without lane keeping assistance system (*No Assistance, NA*) and served as control group.

The second independent variable, the occurrence of *critical driving situations*, was varied within-subjects. Critical driving situations required drivers to react either primarily on the lateral control or on the longitudinal control level in order to avoid collisions with other traffic. Those situations that required drivers to react with respect to lateral control simulated *functional system limits*, because the lane keeping assistance provided by the systems was inadequate in these situations. Drivers had to deviate from the centre of the lane in order to avoid collisions in those situations, and thus had to override the HC system or to ignore the warnings of the LDW system. The critical situations did not represent functional system limits for drivers in the control group (No Assistance, NA) because they were driving without a lane keeping assistance system.

The latency and magnitude of drivers' reactions to the different types of critical situations were used as performance-based measures of their Situation Awareness (SA) when driving with different levels of lane keeping assistance.

The study consisted of two experimental sessions. During the first session, no critical driving situations occurred. Differences in drivers' reliance and in underlying energetic and motivational processes that would be observed between the three groups of drivers in this first session were therefore attributable to the effect of the level of lane keeping assistance.

Critical driving situations occurred however during the second experimental session. It was hypothesised that the occurrence of critical driving situations (partly representing functional system limits) would have a different impact on drivers' reliance and underlying processes for the three levels of lane keeping assistance.

Table 2. Overview about major independent and dependent variables used in the study.

	Manipulation within-subjects: Occurrence of critical driving situations (Session)	
Manipulation between-subjects: Level of lane keeping assistance	No critical situations (Session 1)	Critical situations[a] (Session 2)
	Dependent variables:	
Heading Control (HC)	Reliance & mediating variables	
		SA
Lane Departure Warning (LDW)	Reliance & mediating variables	
		SA
No Assistance (NA)	Visual attention allocation[b] & mediating variables	
		SA

[a] Three of the eight critical situations simulated functional system limits of the HC and LDW system.

[b] Drivers' visual attention allocation strategies between the road scene and the visual display of a secondary task served as a measure for their reliance on the lane keeping assistance systems. The control group's visual attention allocation strategies served as a reference for evaluating the degree of drivers' reliance on the HC and LDW system.

In the following, first research hypotheses referring to the effect of the level of lane keeping assistance on drivers' reliance and underlying processes (when no critical driving situations occurred) will be presented. Thereafter, hypotheses are derived regarding the differential effect of the occurrence of critical driving situations on drivers' reliance and underlying processes for the three level of lane keeping assistance. Last, research hypotheses regarding the effect of the level of lane keeping assistance on drivers' Situation Awareness (SA) are presented.

2.5.3.2 Effects of the level of lane keeping assistance on drivers' reliance and mediating variables

Drivers' lane keeping performance

The assessment of drivers' lane keeping performance served primarily as manipulation check of the variation of the level of lane keeping assistance. Differences in dependent measures may only be attributed to the effects of the level of automation of the lateral control task if the lane keeping assistance systems served their purpose and enhanced drivers' lane keeping performance. The following research hypotheses were derived:

H1: Drivers' lane keeping performance improves as a function of the increasing automation of the lateral control subtask. Lane keeping performance is best for HC drivers and worst for drivers driving without lane keeping assistance (control group). LDW drivers show a medium level of lane keeping performance.

H2: Drivers' lane keeping performance deteriorates during performance of the visual demanding secondary task compared to driving periods without concurrent secondary task performance. This impairment in lane keeping performance due to the concurrent performance of the secondary task is most pronounced for drivers without lane keeping assistance, and least pronounced for HC drivers. LDW drivers show a medium level of impairment in lane keeping performance when concurrently performing the secondary task, compared to driving without secondary task.

Drivers' reliance on the lane keeping assistance systems

Previous studies on operators' reliance on automation yielded evidence that operators are prepared to allocate increasing levels of control to an automated system the higher its reliability. Increasing reliance on automation was furthermore found to be accompanied by a reduction in operators' effort to monitor and verify the automation's actions. A central hypothesis in this study is that the level of automation has a similar effect on reliance as automation reliability. It is assumed that both automation reliability and the level of automation influence reliance by common mechanisms based on the operators' reduced need to intervene in the automation-controlled processes.

Drivers should be prepared to allocate increasing levels of control to the lane keeping assistance systems the more completely they were automating the lateral control driving subtask as a strategy to better deal with the dual-task demands. Drivers' increasing reliance on the lane keeping assistance systems should be reflected in their growing preparedness to allocate their visual attention away from the road scene towards the visual display of a secondary task.

H3: Drivers' reliance on the lane keeping assistance systems increases as a function of the level of automation of the lateral control driving subtask. Drivers rely to a higher extent on the HC system than on the LDW system, indicated by their greater preparedness to allocate their visual attention away from the road scene to the visual display of a concurrent secondary task. Drivers rely to some extent also on the LDW system, indicated by a greater preparedness to allocate their visual attention away from the road scene during concurrent secondary task performance compared to drivers driving without lane keeping assistance.

Mental effort regulation

Differences in drivers' reliance (i.e., in their visual attention allocation strategies) as a function of the level of lane keeping assistance were assumed to be modulated by differences in their mental effort regulation strategies.

The regulation of mental effort is expected to play a major role in influencing drivers' reliance. Theoretical frameworks assume that operators' intention to rely on automation is guided by a utility judgement regarding the costs and benefits associated with the allocation of different levels of control to automation (Dzindolet et al., 2002). Both costs and benefits depend highly on the performance of the automation, i.e. on how efficiently the automation may achieve the operator's goals. The major benefit associated with operators' increasing reliance on highly reliable or high-level automation seems to be the possibility to maintain or even to improve task performance (due to the automation support) while simultaneously limiting the expenditure of mental effort. It was found that operators' reliance on automation varied inversely with the effort they invested in monitoring the automation and verifying its actions (Bagheri & Jamieson, 2004; Bahner et al., 2008; Muir & Moray, 1996). Thus, increasing reliance on automation is associated with operators' less active engagement in the automation-controlled task. It is supposed that operators rely on automation as long as it permits them to maintain an acceptable driving performance while conserving mental effort.

It is assumed that drivers rely to a greater extent on higher levels of lane keeping assistance because they adopt to a greater extent an effort conservation strategy allowing them to maintain primary task performance standards (especially lane keeping performance) while reserving a higher amount of resources for the performance of the secondary task. Drivers' increasing reliance on the lane keeping assistance systems thus should be mirrored by a reduced expenditure of driving-task related effort. Furthermore, higher reliance on the lane keeping assistance systems should be paralleled by a greater tendency to delegate responsibility for maintaining primary task performance standards to the systems.

A higher level of reliance should therefore be associated with lower costs in the form of latent performance decrements in the attempt to maintain performance stability under dual-task conditions. On the contrary, a lower level of reliance on the lane keeping assistance systems (as well as no reliance at all for drivers in the control group) should be accompanied by greater costs as a response to the higher task demands. These costs were assumed to occur in the form of subsidiary task failures and compensatory costs (Fairclough, 2001; Hockey, 1997).

Compensatory costs. Compensatory costs refer to the physiological and affective side effects of human regulatory behaviour in the attempt to protect task performance under high workload conditions.

H4: The magnitude of compensatory costs varies inversely with drivers' reliance on the lane keeping assistance systems. Drivers invest less mental effort in the primary driving task, experience a lower level of mental workload, and feel less frustrated the more they rely on the lane keeping assistance systems. Compensatory costs are most pronounced for drivers in the control condition (driving without lane keeping assistance), and least pronounced for HC drivers.

Subsidiary task failures. Subsidiary task failures refer to the selective impairment of (currently) lower priority task components. Subsidiary task failures were assumed to become evident in drivers' neglect of concurrent secondary task performance as a response to the higher overall task demands. Subsidiary task failures were assumed to be indicated by drivers' longer reaction times and by a decrease in response accuracy in the secondary task.

H5: Drivers' secondary task performance improves as a function of their increasing reliance on the lane keeping assistance systems. HC drivers show the best secondary task performance and drivers in the control condition (driving without lane keeping assistance) show the worst secondary task performance.

Arousal and vigilance

Reliance on automation is assumed to be influenced by operators' subjective state, which can be described by energetic concepts such as vigilance and arousal. Monitoring of automation is considered a resource-intensive vigilance task (Parasuraman, Molloy et al., 1996; Warm et al., 1996). Reliance on automation was found to be associated with operators investing less effort in automation monitoring and verification of its actions. Thus, it can be assumed that operators' tendency to (uncritically) rely on automation increases to the extent that their ability for vigilant system monitoring declines.

Automation may impair vigilance through two mechanisms: First, through resource depletion due to the added cognitive demands associated with automation monitoring; and second, through a low level of arousal caused by a reduction of workload due to automation (see chapter 2.3.5).

At present, there is a lack of empirical studies that systematically investigated the effects of different levels of automation on operators' vigilance, energetic arousal, monitoring behaviour and reliance. It is assumed that increasing levels of automation lead to a decrease in operators' arousal because of the reduced manual control demands, which in turn fosters operators' reliance on automation. Studies on the effects of varying levels of automation reliability provided evidence that operators' energetic arousal varied inversely with automation reliability (Buld & Krüger, 2002). Thus, it is assumed that differences in drivers' reliance on the lane keeping assistance systems are paralleled by differences in drivers' level of arousal. High reliance on the lane keeping assistance systems should associated with lower levels of arousal and reduced vigilance.

H6: Drivers' increasing reliance on the lane keeping assistance systems is accompanied by a reduction in their energetic arousal and vigilance. HC drivers experience the lowest level of arousal, whereas drivers in the control condition (driving without lane keeping assistance) experience the highest level of arousal.

Drivers' trust in the lane keeping assistance systems

In theoretical frameworks, operators' trust in automation has a predominant role in influencing their reliance on automation (e.g., Lee & See, 2004). In general, people rely to a higher extent on automation that they trust. Trust in automation was also found to be negatively correlated with operators' monitoring behaviour. Thus, the more operators trusted an automated system, the less intensely they monitored it. However, not all studies found a positive linear relationship between trust and reliance. This suggests that other factors are also relevant in mediating operators' reliance on automation.

Up to now, trust in automation was mostly investigated as a function of automation reliability. Operators' trust in automation is strongly influenced by the performance of the automation. A decline in automation reliability was generally accompanied by a decline in operators' trust in the automation. The automation's competence (in successful doing what it should do) turned out to a better predictor for operators' level of trust in it than other more process-related aspects of its performance (e.g., the predictability of its behaviour).

Theoretical frameworks make no assumptions about the effect of different degrees of automation on operators' trust. At present it is not clear whether operators trust automation to a higher extent that is more capable in performing a particular task or function, such as lane keeping. In this study, it could be argued that the HC system was more capable or competent in helping drivers to stay within the driving lane than the LDW system by applying continuous steering torques that were designed to guide the vehicle within the centre of the lane. On the contrary, the LDW system did not relieve drivers from manual lateral control, but solely provided warnings when a safety limit was reached (when drivers were just about crossing the lane markings). Thus, it was hypothesised that drivers would trust the HC system to a higher extent than the LDW system. However, this effect was thought to be smaller in magnitude than the effect of varying levels of automation reliability on operators' trust. Also, it was assumed that drivers' reliance on the lane keeping assistance systems was influenced to a higher extent by motivational and energetic processes than by drivers' trust in the systems.

H7: Drivers' increasing reliance on the lane keeping assistance systems is accompanied by a higher level of trust in the lane keeping assistance systems. Drivers trust the HC system more than the LDW system.

2.5.3.3 Effects of drivers' experience of functional system limits on their reliance and mediating variables

Drivers' reliance on the lane keeping assistance systems

Drivers' experience of functional system limits in the second experimental session should have a negative effect on their reliance on the lane keeping assistance systems. It was expected that the experience of functional system limits has a similar effect on reliance as incidences of system failures in previous studies. It was found that, in general, operators relied to a lesser extent on automation after having experienced that it may fail. The decline in reliance should result from drivers' perception that the lane keeping assistance systems were not sufficiently effective for maintaining an adequate driving performance under the changing task conditions. Recent theoretical frameworks assume that reliance is not a static behaviour such that operators allocate a predefined level of control to automation based on its performance or capabilities, but that reliance is sensitive to momentary changes in task demands and in automation performance (Lee & See, 2004).

The experience that the lane keeping assistance systems' support was inadequate under some circumstances (in situations representing functional system limits) should cause drivers to rely less on the lane keeping assistance systems in order to maintain acceptable standards of (driving) performance and safety when faced with unexpected hazardous driving situations. The decline in reliance should primarily result from changes in drivers' strategies for mental effort investment, but it should also be mediated by changes in energetic processes and drivers' attitudes towards the lane keeping assistance systems.

As a response to the occasional experience of functional system limits and the unexpected encountering of critical driving situations drivers were expected to be less prepared to allocate their visual attention away from the road scene (indicative for their lower reliance on the lane keeping assistance systems). This effect should also occur for drivers driving without lane keeping assistance as a result of a shift in their effort management and attention allocation strategies in adaptation to the increase in primary task demands. However, it was hypothesised that the shift in resource allocation strategies (and its effect on reliance) should be most pronounced for HC drivers, and less pronounced for LDW drivers. Thus, it was predicted that the decline in drivers' reliance on the lane keeping assistance systems should be more pronounced for the HC drivers than for the LDW drivers.

This was expected because of the greater intervention of the HC system in drivers' lateral control processes and the herewith associated larger negative consequences on performance and safety when the support of the lane keeping assistance systems was inadequate. Furthermore, those negative consequences were to a higher degree inevitable with the HC system. Because drivers had no possibility to switch the lane keeping assistance systems off (despite of using the turn indicator in order to signal an intentional departure from the driving lane), even a low level of reliance on the HC system was associated with the allocation of a substantial level of control to the system. Drivers had to actively override the HC system in situations that demanded their reaction on the lateral control level in order to avoid a collision. On the contrary, drivers could 'simply' ignore the warnings of the LDW system in situations requiring responses on the lateral control level, but were not affected in their driving behaviour.

It was expected that the hypothesised decline in drivers' reliance due to the experience of functional system limits is mirrored by changes in motivational and energetic processes as well as by changes in drivers' attitudes towards the lane keeping assistance systems. It must be kept in mind that drivers did neither know *that* they would encounter critical driving situations, nor *when* and *how many* of them would occur. That means that changes in drivers' reliance and underlying processes were expected to result from drivers' *perception* of increased task demands based on the *anticipation* of their need to potentially intervene in some circumstances. Also drivers driving without lane keeping assistance were expected to change their attention management and resource allocation strategies in response to the occurrence of critical driving situations. However, it was assumed that the shift in strategies would be smaller in magnitude for them.

H8: Drivers are less prepared to allocate their visual attention from the road scene towards the in-vehicle display of a secondary task as a response to the unexpected occurrence of critical driving situations (partly representing functional system limits). This shift in attention allocation strategies is larger the more prepared drivers were to allocate their visual attention away from the road scene before they encountered critical driving situations. Thus, the shift in attention allocation strategies is most pronounced for HC drivers, and less pronounced for LDW drivers, indicating a greater decline in drivers' reliance on the HC system than on the LDW system. The shift in visual attention allocation strategies is least pronounced for drivers driving without lane keeping assistance.

H9: HC drivers are least prepared to allocate their visual attention from the road scene as a response to the unexpected occurrence of critical driving situations (partly representing functional system limits). LDW drivers and drivers driving without lane keeping assistance (control group) are more prepared than HC drivers to allocate their visual attention away from the road scene.

Mental effort regulation

It was predicted that the occurrence of critical driving situations leads to an increase in task demands, and that drivers react to this increase in task demands by changing their strategies for mental effort regulation. Specifically, it was hypothesised that drivers invest more effort in the driving task in order to adapt to the increasing task demands. This shift to more resource-intensive performance strategies was assumed to result in greater costs occurring in the form of latent performance decrements under dual-task conditions (Hockey, 1997).

The shift in mental effort regulation strategies should be most pronounced for HC drivers because they did not only have to react appropriately to the critical situations (just like drivers in the two other experimental groups), but also to actively override the HC system in order to avoid collisions with other traffic. Thus, the occurrence of functional system limits should lead HC drivers to invest more effort in the driving task and in monitoring the HC system's actions in order to be able to circumvent the detrimental effects of the HC system on driving performance in situations representing functional system limits. The hypothesised larger decline in drivers' reliance on the HC system in response to the occurrence of critical driving situations should thus be mirrored by a larger increase in their effort for maintaining an acceptable driving performance under the unexpected occurrence of critical situations.

Drivers' shift to more resource-intensive performance strategies (in relation with a decline in their reliance on the lane keeping assistance systems) was hypothesised to be accompanied by larger costs in the form of latent performance decrements. Those should become evident as compensatory costs and subsidiary task failures.

H10: The increase in task demands due to the unexpected occurrence of critical driving situations demanding for drivers' intervention results in an increase in compensatory costs associated with drivers' attempts to maintain performance stability. In response to the unexpected occurrence of critical driving situations drivers invest more mental effort in the driving task, experience a higher level of mental workload, and feel more frustrated. The increase in compensatory costs is more pronounced the more drivers adopted a mental effort conservation strategy prior to the occurrence of critical driving situations. Thus, the increase in compensatory costs is most pronounced for HC drivers, and least pronounced for drivers driving without lane keeping assistance.

H11: The magnitude of compensatory costs is larger for HC drivers than for LDW drivers and drivers driving without lane keeping assistance in response to their encountering of critical driving situations. HC drivers invest most mental effort in the driving task, experience the highest level of mental workload, and feel most frustrated in response to the unexpected occurrence of critical driving situations (partly representing functional system limits). LDW drivers and drivers driving without lane keeping assistance invest less mental effort in the driving task, experience a lower level of mental workload, and feel less frustrated than HC drivers in response to the unexpected occurrence of critical driving situations.

H12: In response to the occurrence of critical driving situations (partly representing functional system limits) drivers' secondary task performance deteriorates. The impairment in secondary task performance is more pronounced the more drivers adopted a mental effort conservation strategy prior to the occurrence of critical driving situations. Thus, the impairment in secondary task performance is most pronounced for HC drivers, and least pronounced for drivers driving without lane keeping assistance.

H13: HC drivers show the worst performance in the secondary task in response to the unexpected occurrence of critical driving situations (partly representing functional system limits), whereas LDW drivers and drivers driving without lane keeping assistance show a better secondary task performance than HC drivers.

Arousal and vigilance

It was hypothesised that the increase in task demands due to the occurrence of critical driving situations would have a differential impact on drivers' arousal and vigilance based on their mental effort regulation strategies in the first experimental session. It was expected that HC drivers were to a higher extent able to reserve attentional resources during the first experimental session (where no critical situations occurred) compared to LDW drivers and drivers driving without lane keeping assistance because they were hypothesised to adopt to a greater extent an effort conservation strategy in the first experimental session. Thus, HC drivers were assumed to react to the increase in task demands caused by the unexpected occurrence of critical driving situations by a mobilisation of "reserve" attentional resources that in turn should lead to an increase in arousal and vigilance. It was therefore assumed that the expected decline in drivers' reliance on the HC system in response to the unexpected occurrence of critical driving situations is accompanied by an increase in HC drivers' arousal and vigilance.

On the contrary, drivers driving without lane keeping assistance were supposed to have already invested a higher level of mental effort in the driving task during the first experimental session (because they could not rely on a lane keeping assistance system). A further increase in task demands caused by the occurrence of critical driving situations was assumed to lead to a further depletion of attentional resources that in turn should lead to a decline in arousal and vigilance.

LDW drivers were supposed to have invested more mental effort in driving during the first experimental session than HC drivers, but less mental effort than drivers in the control group. Thus, it was expected that their level of arousal should not change that much in response to the increasing task demands.

> *H14: The increase in task demands due to the unexpected occurrence of critical driving situations leads to an increase in arousal and vigilance for HC drivers (because of their mobilisation of reserve attentional resources), and to a decline in arousal and vigilance for drivers driving without lane keeping assistance (because of a depletion of attentional resources). LDW drivers' level of arousal does not change in response to the occurrence of critical driving situations.*

Drivers' trust in the lane keeping assistance systems

Previous studies showed that operators' trust in (and as a result, their reliance on) automation declines as a response to automation failures. A similar effect (though maybe less pronounced) was predicted to result from drivers' experience of functional system limits. Drivers' trust was hypothesised to decline to a higher extent in response to functional limits of the HC system, and to a lower extent in response to functional limits of the LDW system, because of the more adverse consequences of functional limits of the HC system on driving performance and safety. Thus, the predicted stronger decline in HC drivers' reliance in response to their experience of functional system limits should be accompanied by a larger decline in their level of trust in the HC system compared to LDW drivers.

> *H15: Drivers' trust in the lane keeping assistance systems declines as a response to the occurrence of functional system limits. The decline in trust is more pronounced for drivers driving with the HC system than for drivers driving with the LDW system.*

2.5.3.4 Effects of the level of lane keeping assistance on drivers' situation awareness

Drivers' reactions to the critical driving situations occurring during the second experimental session were taken as a measure of their Situation Awareness (SA) when driving with different levels of lane keeping assistance.

It was predicted that automation interrupts the successful alternation of top-down and bottom-up attentional processes which are necessary to build up and maintain a comprehensive and current mental representation of the situation. It was expected that with increasing automation of the lateral control subtask, drivers allocate fewer attentional resources to the processing of visual cues relevant for lateral control of the vehicle. This reduced processing of information relevant for lateral control should result in failures in bottom-up control of attention in driving situations that required drivers to react by adjustments in lateral control of the vehicle (e.g., by drawing aside) in order to avoid a collision with other traffic participants.

In order to test this hypothesis, different categories of critical situations were defined that differed (a) to whether they required drivers to react primarily on the lateral or on the longitudinal control level in order to avoid a collision, (b) with respect to the time horizon available for drivers to perceive and interpret environmental cues and to plan ahead their response, and (c) to the demands they placed on higher-level cognitive processes, i.e. to whether they required a high or a low level of anticipation (see Table 1 on page 82).

The different categories of situations were designed in order to investigate the hypothesised changes in the depth of processing of visual information either relevant for lateral or longitudinal control of the vehicle. Level of assistance effects were only expected in situations that required drivers to react on the lateral control level and not in situations requiring drivers to react primarily on the longitudinal control level. Furthermore, differences in the depth of processing of information relevant for lateral control should become especially apparent in "unforeseeable" situations that did drivers not leave much time to perceive and interpret visual information and to prepare their reaction, as well as in situations that required a higher amount of anticipation (higher-level cognitive processing) of the behaviour of other traffic participants and of the future development of the situation. Failures in bottom-up attentional processes were assumed to result in drivers' delayed and more abrupt reactions to these situations. On the contrary, drivers should be able to compensate for failures in bottom-up attentional control in driving situations that leave them more time to perceive and interpret environmental cues and to plan ahead their reaction. Thus, no effect of the level of lane keeping assistance was expected in driving situations that were categorised as 'lateral foreseeable'.

The following research hypotheses were derived:

H16: Drivers do not differ in their reaction (in terms of the starting point of their initial reaction and in terms of the abruptness of their reaction) in response to driving situations that require them to react primarily on the longitudinal control level.

H17: HC drivers react later and more abruptly than LDW drivers and drivers in the (manual) control group to driving situations that require them to react on the lateral control level and that require a high amount of anticipation of the behaviour of other traffic participants ("lateral anticipatory" situations). Also, HC drivers react later and more abruptly than LDW drivers and drivers in the (manual) control group to driving situations that require them to react on the lateral control level and that do not leave them much time for the processing of the relevant visual cues and for response planning ("lateral unforeseeable" situations). Drivers in the control group show the earliest and smoothest reactions to these situations, whereas LDW drivers react later and more abruptly to these situations than drivers in the (manual) control group, but earlier and less abruptly than HC drivers.

H18: Drivers do not differ in their reaction (in terms of the starting point of their initial reaction and in terms of the abruptness of their reaction) to driving situations that require them to react primarily on the lateral control level, but that leave them a comparably high amount of time to perceive and interpret visual information and to plan ahead their response ("lateral foreseeable" situations).

3 METHODS

3.1 Participants

Forty-five drivers (25 male, 20 female) aged between 25 and 46 years (*M* = 33.7; *SD* = 6.4) took part in this study. They were recruited from a pool of Swedish drivers who had reported their interest in participating in driving simulator studies at the Swedish National Road and Transport Research Institute (VTI) in Linköping. Thirty-six drivers (80%) had been driving in the VTI driving simulator before. The other nine drivers had no previous experiences with driving in the VTI driving simulator. The participants received 400 Swedish crowns (about 42 Euros) as monetary compensation for their participation in the study.

Drivers were in possession of a valid driving licence for an average of 14.3 years (*SD* = 7.4). Their driving experience was assessed in six categories of their total distance driven, ranging from 0 to 10,000 km (first category) to more than 150,000 km (sixth category, see Figure 3). Drivers had been driving a minimum of 10,001 to 30,000 km before they took part in the study. Almost half of them (40%) had been driving more than 150,000 km. In the past 12 months, participants had been driving between 1,500 km and 55,000 km (*M* = 14,283.3; *SD* = 10,251.3).

In order to assess drivers' visual attention allocation strategies, their eye movements were recorded by the help of an eye-tracking system. In order to ensure a high tracking quality, participation in the study was restricted to drivers with normal visual acuity or drivers who wore contact lenses. Spectacle wearers were excluded from participation in the study.

Participants were matched for gender and as much as possible for age and were then randomly assigned to one of three experimental groups representing different levels of lane keeping assistance. Participants who drove with a *Heading Control (HC)* system (representing a high level of lane keeping assistance) or with a *Lane Departure Warning (LDW)* system (representing a low level of lane keeping assistance) had no previous experiences in driving with these systems. Ten participants (two HC drivers, four LDW drivers, and four drivers in the *No Assistance* group) had previous experiences in driving with an Adaptive Cruise Control (ACC) system. Five of them reported that they had been driving "a few times" with an ACC system, whereas the other 5 participants reported that they had been driving "often" with an ACC system.

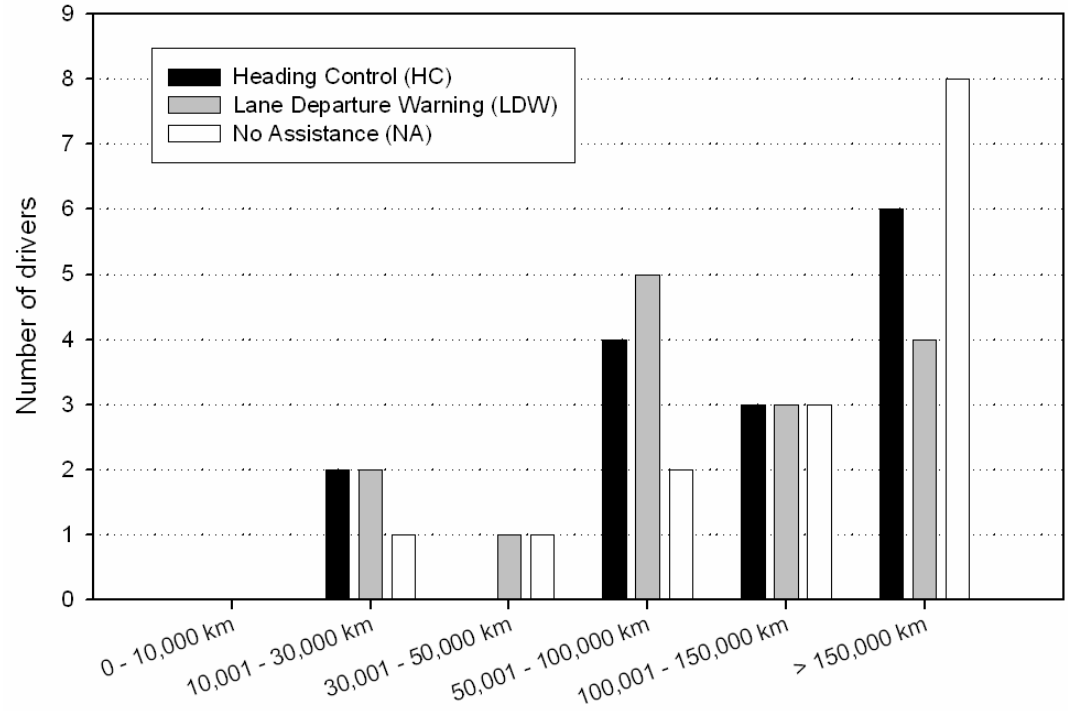


Figure 3. Participants' driving experience across the three experimental groups (representing different levels of lane keeping assistance) as assessed by six categories of total distance driven [km] after they obtained their driving licence.

3.2 Driving Simulator

The advanced moving-base driving simulator III at VTI in Linköping, Sweden, was used in this study (see Figure 4a-f)[2]. It is built around a real vehicle chassis (Volvo 850 with manual gearbox) and has a sophisticated motion system simulating acceleration in three dimensions through roll, pitch and external linear motion in either lateral or longitudinal direction. Because the focus in this study was on the effects of different levels of automation of the lateral control driving subtask, the upper section of the simulator (chassis and screen) was turned 90° in order to simulate lateral forces while driving and to create realistic impressions of the actions of the lane keeping assistance systems (especially of the HC system). Together with a vibration table, a computerized vehicle model, a wide-angle visual system and a sound system the simulator generates very realistic experiences while driving. The surroundings are simulated and displayed

---

[2] This driving simulator study was conducted in the framework of the European Network of Excellence HUMANIST (HUMAN centred design for Information Society Technologies) and was supported by a 3-years PhD grant and by a € 78,000 EC-grant for infrastructure-sharing that partly covered costs for simulator hire, personal (programming the simulator scenarios, experimental leaders), equipment, data logging, pilot testing, participant fees, and for the PhD student’s stay abroad (travel expenses).

to the driver via one main screen with a 120° horizontal field of view 2.5 m in front of the driver, and three rear-view mirrors (monitors). The computer graphics have a very short time delay of less then 40 ms.

Participants' driving behaviour was recorded with a sampling rate of 200 Hz. Table 3 gives an overview about the technical data of the VTI driving simulator.

Table 3. Technical data of the VTI moving-base driving simulator III.

Motion system	
Pitch angle	-9 degrees to +14 degrees
Roll angle	± 24 degrees
External linear motion	
Maximum amplitude	± 3.75 m
Maximum speed	± 4.0 m/s
Maximum acceleration	± 0.8 g
Vibration table	
Vertical movement	± 6.0 cm
Longitudinal movement	± 6.0 cm
Roll angle	± 6 degrees
Pitch angle	± 3 degrees

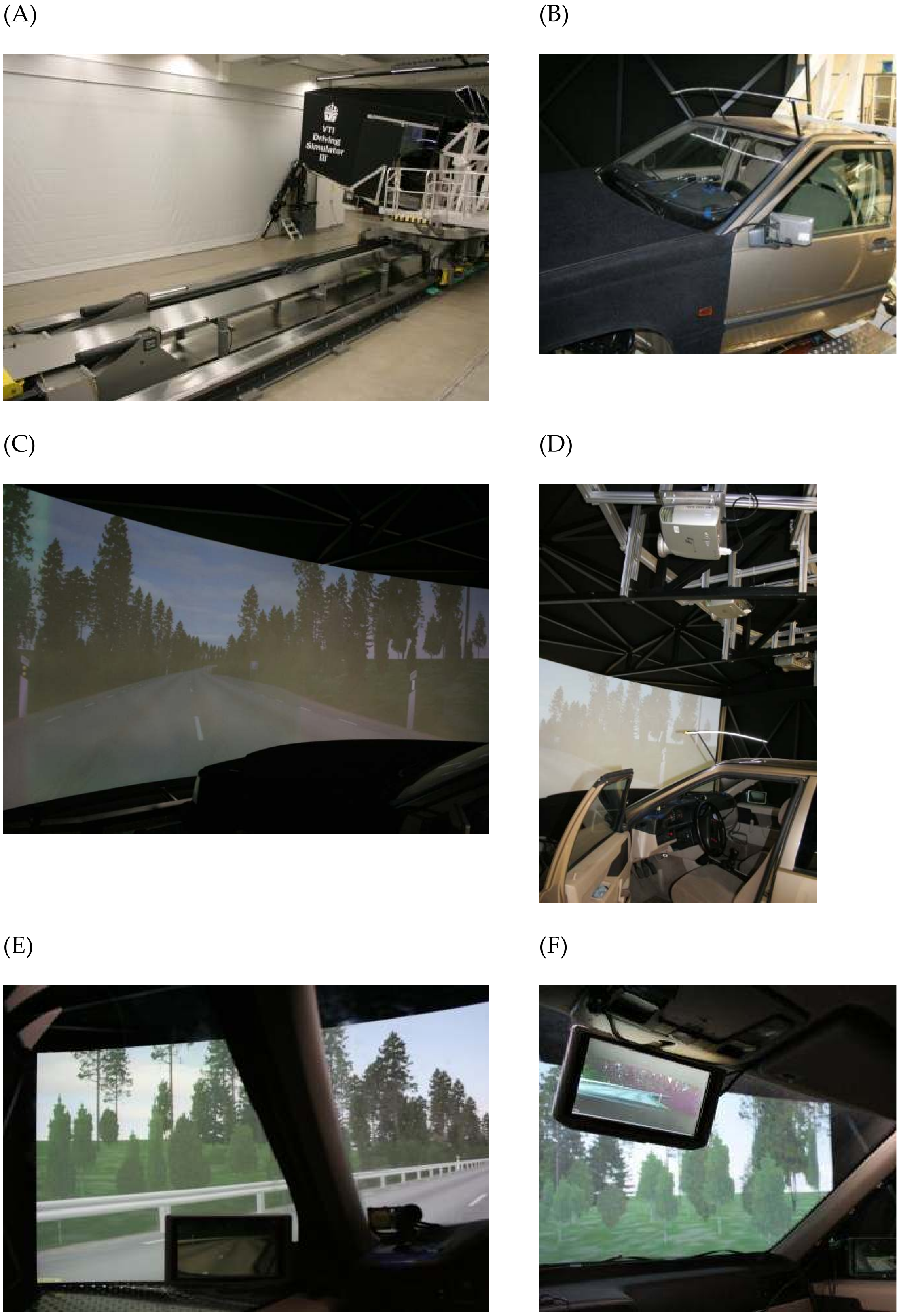


Figure 4. (a) VTI moving-base driving simulator. (b) simulator vehicle chassis (Volvo 850). (c) wide-angle visual screen with 120° horizontal field of view. (d) visual projection system. (e) left outside mirror (monitor). (f) rear mirror (monitor) and right outside mirror (monitor).

3.3 Driving Task

3.3.1 Experimental route

Participants drove on a rural road consisting of two 3.5 m-wide opposing lanes and lay-bys on each side of the road. The road led through a wooden area and was a virtual simulation of a real existing Swedish road near Linköping. The road was delimited by reflexion posts and/or crash barriers on each side that were placed 0.5 m outside the outer lane markings (continuous white lines), yielding a total road width of 8 m. The road's maximum gradient was 3.5%, and its minimum gradient was -5.71% (*M* = -0.6%). The road consisted mainly of straight road sections, but also included gentle curves that had a minimum radius of 968 m (right curves) up to 1200 m (left curves). There was no other traffic in the same lane (except during some of the critical driving situations that occurred during the second experimental session), and light traffic in the opposite lane with an average density of 3.25 vehicles/km. The posted speed limit was 70 km/h.

Figure 5 shows an exemplary view of the road scene.

The road had a length of 10.8 km and was driven six times by each driver, yielding a total length of 64.8 km. Drivers drove the road one time during the practice session (without secondary task, without critical driving situations), two times during the first experimental session (with secondary task, without critical driving situations), and three times during the second experimental session (with secondary task, with critical driving situations).

Figure 5. Prototypical view of the experimental route.

3.3.2 Lane keeping assistance systems

Two lane keeping assistance systems were used in this study in order to implement a high and a low level of lane keeping assistance. A high level of lane keeping assistance was realised by a *Heading Control (HC)* system. A low level of lane keeping assistance was realised by a *Lane Departure Warning (LDW)* system.

The lane keeping assistance systems were designed to prevent drivers from unintentional departures from the driving lane caused by the visual distraction from their concurrent performance of the secondary task. The two lane keeping assistance systems differed with respect to the degree of continuousness with which they provided drivers with information about their lateral position in the driving lane.

The LDW system warned drivers via a binary (on/off) non-directional haptic warning on the steering wheel of a near-by crossing of the lane boundaries. Thus, the LDW system did not relieve drivers from lateral control of the vehicle, but provided them with additional information when they were just about violating a specific safety-limit (departure from the driving lane). The HC system provided drivers with continuous haptic feedback on the steering wheel about their lateral position in the driving lane that varied according to the magnitude of the vehicle's lateral displacement from the centre of the lane. The HC system additionally relieved drivers from some manual control demands by applying corrective counter forces on the steering wheel designed to return the simulator vehicle back towards the centre of the driving lane.

Both systems could be switched off by drivers activating the turn indicator in order to signal their intention to deviate from the driving lane (e.g., in order to perform a lane change). The HC and the LDW system were 100% reliable during the study. However, during the second experimental session drivers encountered critical driving situations that partly required them to react on the lateral control level (e.g., by drawing aside) in order to avoid collisions with other traffic. Those situations were designed to represent functional limits of the lane keeping assistance systems, because the systems were not designed to detect those situations and hence, offered inadequate support under these circumstances. Thus, drivers had to react accordingly to these critical driving situations by either switching off the lane keeping assistance systems, or by overriding the HC system's steering torques and ignoring the LDW system's warnings, respectively.

3.3.2.1 Functionality of the Heading Control system

The HC system was designed to keep drivers within the driving lane by applying counter forces on the steering wheel towards the centre of the lane depending on the vehicle's lateral displacement from the centre of the lane. The torques generated by the HC system were strong enough to return the simulator vehicle back close to the centre of the lane provided that drivers did not apply counter pressure on the steering wheel in the opposite direction. The HC system's maximum torque was limited to 2.4 Nm.

The magnitude of the HC system's steering torques varied according to the state of a "controller" that could reach values between 0 and 1 dependent on the vehicle's lateral deviation from the centre of the lane *(dy)*. When drivers drove perfectly in the centre of the lane, the steering wheel torque was computed by the vehicle model, comparable to the steering column friction of a real driving car. As soon as the vehicle deviated from the centre of the lane, the steering wheel torque was generated to different portions by the vehicle model and by the controller. Thus, different states of the controller *(k)* represent different ratios of the vehicle model's and the controller's generation of the steering torque (see Figure 6). This was done in order to obtain a smooth transition between the two control modes.

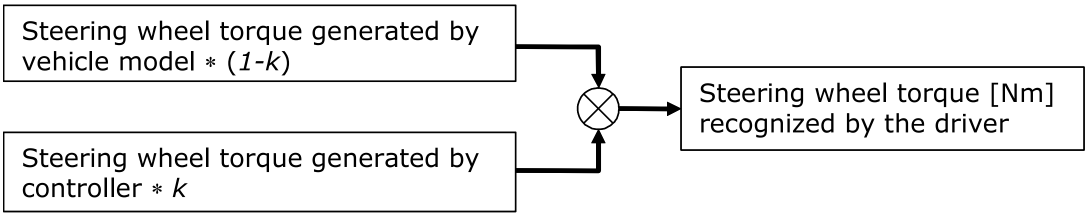


Figure 6. Computation of the actual steering torque as a mixture of the vehicle model's and the conroller's generated steering torque as a function of the state of the controller (k).

The larger the vehicle's actual deviation from the centre of the lane, the larger became the steering wheel torque generated by the controller. If the vehicle's lateral position was close to the centre of the lane, than *k* was close to 0 and the steering wheel torque experienced by the driver was mainly generated from the vehicle model. When *dy* became equal to ± 0.6 m (corresponding to a lateral distance of 0.42 m between the centre of the vehicle's front tires and the lane markings), *k* reached its maximum value of 1. In this case, the normal steering wheel torque from the vehicle model was disconnected, and the steering torque of the controller reached its maximum force of 2.4 Nm. If *dy* exceeded ± 0.6 m, *k* was set to 1.

The state of the controller was computed as

$$for\ -0.6 \le dy \ge 0.6; \qquad k = \left\{ 0.5 \times \left[1 - \cos\left(\frac{pi \times |dy|}{0.6} \right) \right] \right\}^2 \tag{1}$$

Figure 7 shows the values of *k* for different values of *dy*.

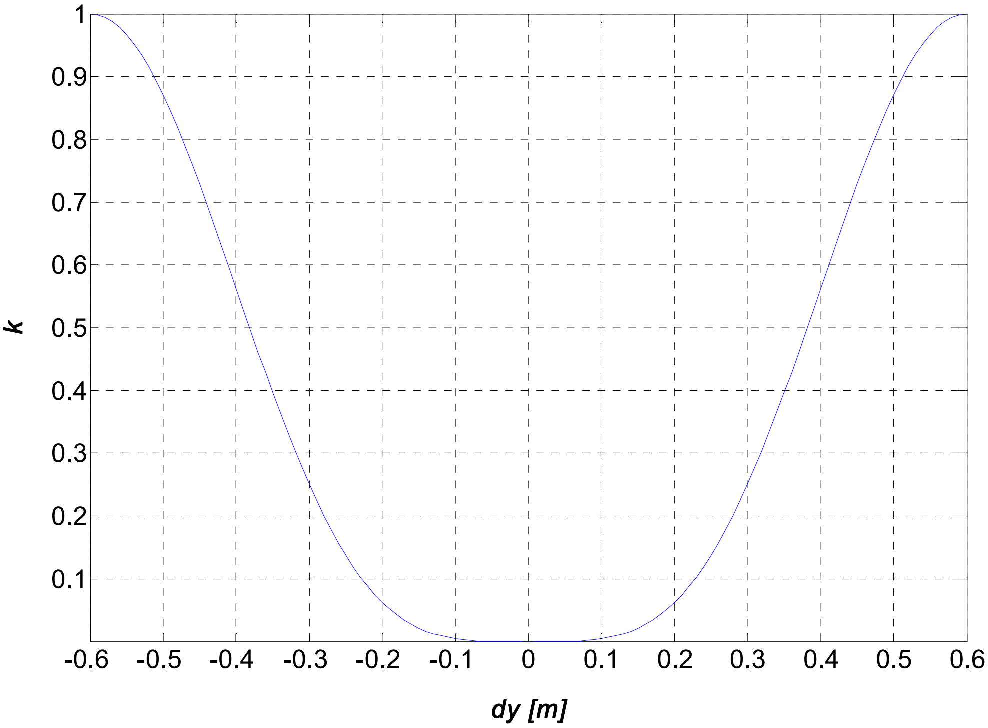


Figure 7. State of the controller *(k)* generating the steering torques of the HC system as a function of the vehicle's lateral deviation from the centre of the lane *(dy)*. If *k* is close to 0, the steering torqueses experienced by the driver resemble those of a normal vehicle. If *k* equals 1, the vehicle model is disconnected and the steering torqueses (maximum 2.4 Nm) are solely generated by the controller.

The HC system could be switched off by drivers through activation of the turn indicator (in order to signal an intentional departure from the driving lane). When drivers activated the indicator, *k* was set to 0 within a time range of 1 s. The controller remained disconnected until *dy* returned to ≤ ± 0.9 m (corresponding to a lateral distance of 0.12 m between the centre of the vehicle’s front tires and the lane markings) for a duration of more than 2 s. After this time, the controller became active again according to Equation (1).

A logic was implemented that automatically disconnected the controller within a range of ± 0.25 m (in order to ensure a transition of *k* from 1 to 0) if *dy* became larger than ± 1.3 m (corresponding to a 0.27 m line crossing). The controller was then reconnected based on a time function (within 1 s after *dy* became ≤ ± 1.3 m). This logic was implemented in order to realise a normal steering behaviour of the simulator vehicle in the event that drivers forgot to activate the turn indicator, provided that the lateral deviation from the lane was large enough to assume drivers’ intentional departure from the driving lane. The above mentioned critical distances were chosen based on pilot tests.

Furthermore, the controller was always disconnected if the vehicle’s velocity was lower than 10 m/s.

3.3.2.2 Functionality of the Lane Departure Warning system

The LDW system was designed to support drivers in lane keeping by providing them with a haptic feedback on the steering wheel (vibration) when they were about to cross the left or the right lane marking. Thus, the LDW system warned drivers of an unintentional lane departure; however, drivers were still responsible to determine the appropriate action (steer to the right or to the left) and to implement this corrective steering action. The LDW system therefore offered a lower level of lane keeping assistance than the HC system.

When the simulator vehicle's lateral distance from the centre of the lane (*dy)* became larger than ± 0.6 m (corresponding to a lateral distance of 0.42 m between the centre of the vehicle's front tires and the lane markings) a steering wheel torque was applied in order to simulate low-level vibrations of the steering wheel. Thus, the critical lateral deviation for triggering a lane departure warning equalled the deviation at which the HC system generated its maximum steering torque of 2.4 Nm. The lane departure warning was a sinus signal with a frequency of 35 Hz. The duration of the warning was 2 s unless the driver returned to a lateral distance of less than ± 0.6 m from the centre of the lane within this period of time. In this case, the warning was switched off as soon as the vehicle's lateral deviation fell below this limit. Drivers could be provided with a new warning as soon as *dy* exceeded again the critical range of ± 0.6 m around the centre of the driving lane.

Like the HC system, the lane departure warnings could be suppressed by drivers through activation of the turn indicator (in order to signal their intention to deviate from the driving lane) as long as *dy* was < ± 0.6 m. The LDW system was automatically switched on again as soon as drivers had returned to a distance of < ± 0.6 m around the centre of the lane and had switched off the turn indicator.

3.4 Secondary Task

Concurrently to driving, participants performed a visually demanding secondary task. The secondary task served two major purposes: First, it should provoke variance in lateral control of the vehicle in order to enable drivers to experience the actions and behaviour of the lane keeping assistance systems (and to foster their reliance on them). Second, drivers' strategies of allocating their visual attention between the road scene and the visual display of the secondary task were used as a measure for their reliance on the lane keeping assistance systems.

The secondary task used in this study was a visual search task of medium complexity developed in the European HASTE project (Carsten et al., 2005; Engström et al., 2005). It was presented on a 7″ TFT touch screen mounted in the centre console of the vehicle (see Figure 8). The visual display was positioned 26 cm to the right of the steering wheel (corresponding to a horizontal eccentricity of about 22° to 37° visual angle to the right of the driver's straight-ahead line of sight), and about 10 cm below the top of the dashboard (corresponding to a vertical eccentricity of about 13° to 22° visual angle below the driver's straight-ahead line of sight).

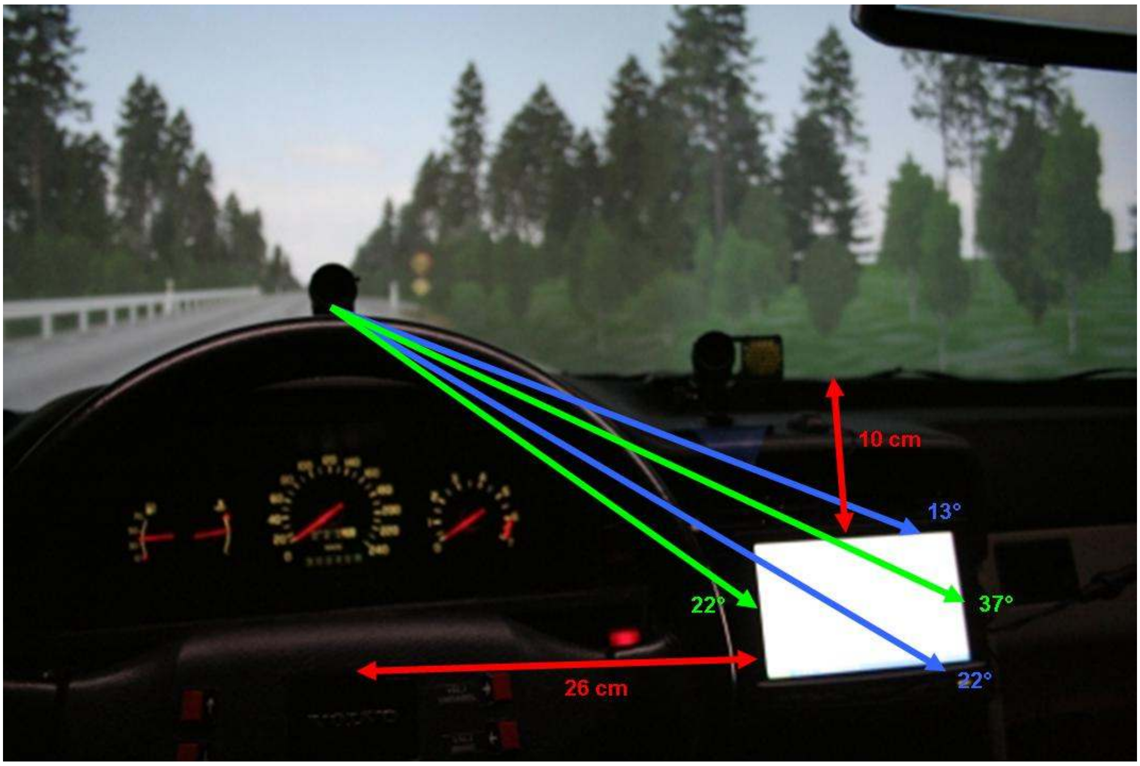


Figure 8. Position of the secondary task's visual display (touch screen) in the centre console of the vehicle.

The secondary task, called hereafter "Arrows Task", consisted of a series of displays representing different configuration of 16 arrows that pointed in different directions. The 16 arrows were organised in 4 rows and 4 columns. For each display, drivers had to decide whether there was an arrow pointing upward present, and to respond accordingly by pressing the "yes" or the "no" ("ja" and "nej" in Swedish) button in the upper or in the lower part of the display, respectively (see Figure 9). The probability of occurrence of an upward-facing arrow was 50%.

Each Arrows Task lasted about 30 s. The number of displays presented in each Arrows Task varied dependent on how fast drivers responded to the single displays. The maximum presentation time for each display was 5 s (time-out). If drivers did not respond to a display within this period of time, this was recorded as a missing response, and a new display was presented again after 1 s.

The presentation of each new display was signalled by a sound. After responding to a display, drivers were provided with a visual and auditory performance feedback. The visual feedback consisted of a text message appearing on the centre of the screen reading either "correct response", "false response", or "missing response". The auditory feedback consisted of two sounds; one for correct responses ("pling") and one for false responses ("toot"). The performance feedback was implemented in order to ensure that drivers took the secondary task seriously and thus, to enable a continuous and reliable measurement of drivers’ visual attention allocation strategies. The time between drivers’ responses to one display and the presentation of a new display was sampled from a uniform distribution ranging from a minimum time delay of 1.22 s and a maximum delay of 4.22 s.

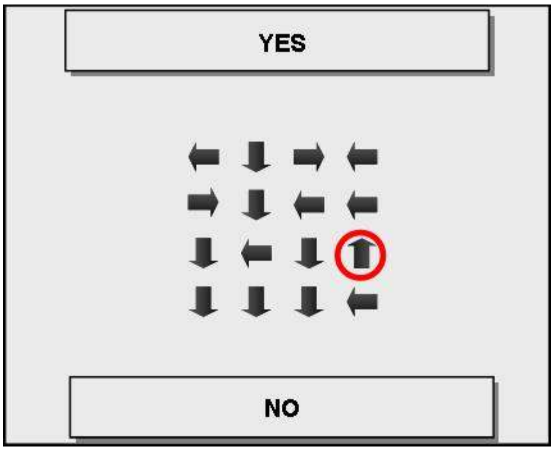


Figure 9. The Arrows Task. It consisted of a series of displays presenting different configurations of 16 arrows pointing in different directions. For each display, drivers had to decide whether there was an arrow pointing upward present (for illustration purposes circled red), and to respond accordingly by pressing the "yes" or the "no" button.

The Arrows Task started at certain road marks which were the same for each driver in each condition. During the first experimental session, drivers performed 19 Arrows Tasks, and during the second experimental session drivers performed 25 Arrows Tasks. Drivers were instructed to perform the Arrows Task as well and as fast as possible while driving, but without jeopardising traffic safety. Dependent measures were drivers’ response times to the Arrows Task displays, and response accuracy.

3.5 Critical Driving Situations

During the second experimental session, drivers encountered eight critical driving situations. Those situations occurred *unexpected* for the drivers and served two main purposes: First, they were designed to assess the effect of the level of lane keeping assistance on drivers’ situation awareness; and second, they were used to investigate changes in drivers’ reliance on the lane keeping assistance systems as a function of their experience of functional system limits.

In order to assess automation-induced changes in drivers’ situation awareness, different types of critical situations were designed. Those different categories served to investigate the hypothesised differences in the depth of drivers’ processing of environmental cues that were either relevant or irrelevant for the driving subtask under (partly) automated control (lateral control) and the driving subtask still under manual control (longitudinal control). The relative point in time of drivers' initial reaction to the different types of critical driving situations and the abruptness of their reaction (initiated driving manoeuvre) served as the major dependent variables in order to assess drivers' SA. Drivers' delayed and more abrupt reactions to situations that required a reaction with respect to either lateral or longitudinal control was assumed to be indicative for an impairment of the cognitive processes necessary to instantiate (activate) information relevant for either lateral or longitudinal control in the actual situation model.

The critical driving situations always occurred *while drivers were concurrently performing the Arrows Task*. Although this hazarded the consequence that the effect of the level of lane keeping assistance on drivers' SA might be confounded by the effect of drivers' visual distraction due to the Arrows Task, it was expected that there would be a greater chance for observing an effect of the level of lane keeping assistance on SA under resource-intensive dual-task conditions than under single-task conditions (driving alone). It is widely acknowledged that the process of building up and maintaining a current and comprehensive mental representation of the actual situation is a resource-intensive process (e.g., Baumann & Krems, 2007). It was supposed that drivers could use eventual reserve mental capacity for compensating for automation-induced impairments in SA under single task conditions.

Table 4 gives an overview about the different situation categories. Seven situations were "safety-critical" meaning that a missing driver reaction in those situations could have potentially resulted in a collision with other traffic (see upper part of Table 4). One situation was categorised as "not safety-critical". In this situation, drivers' tendency to follow the erroneous behaviour of a lead vehicle ("car-following", see lower part of Table 4) was used as a measure of their SA.

The *seven safety-critical* situations differed (a) to whether they required drivers to react primarily on the lateral or on the longitudinal control level in order to avoid a collision, (b) with respect to the "time horizon" available for perception of the relevant visual cues, response planning and execution, and (c) to the demands they placed on higher-level cognitive processes, i.e. to whether they required a high or a low level of anticipation. Effects of the level of assistance on drivers' SA were expected to become apparent only in two situations that required drivers to react with respect to lateral control: in the lateral "unforeseeable" situation, and in the lateral "anticipatory" situation. The critical situations demanding for drivers' reaction with respect to lateral control simulated functional limits of the lane keeping assistance systems because the systems provided inadequate support in those situations.

Table 4. Types of critical driving situations used to investigate content-specific changes in drivers' situation awareness when driving with different levels of lane keeping assistance. Situation categories were realised by specific driving “Events” (labelled alphabetically) that occurred unexpected for drivers in the second experimental session. Numbers in brackets refer to the number of critical situations occurring in the respective category.

Safety-critical (7)		
	Time horizon available	
Anticipation demands	small	large
high	--	“anticipatory” lateral *(Event G)* longitudinal *(Event F, Event E)*
low	“unforeseeable” lateral *(Event C)* longitudinal *(Event D)*	“foreseeable” lateral *(Event A)* longitudinal *(Event B)*
Not safety-critical (1)		
“car-following” longitudinal *(Event H)*		

All critical situations occurred during the second experimental session. This session consisted of three consecutive laps of the experimental route, yielding a total length of 32.4 km (corresponding to 20 to 25 min driving time). During each of the three laps, drivers encountered two up to three critical situations. The order of the three laps was counterbalanced across participants. However, each lap consisted always of the same critical situations that occurred in the same order. Figure 10 shows the assignment of the different types of critical situations to the three laps that drivers drove during the second experimental session.

Several critical situations started or ended with other vehicles pulling in or out from the lay-bys on both sides of the road. In order to avoid that drivers could infer that “something will happen” each time when there were other vehicles standing in the lay-bys, some “dummy” vehicles were introduced that parked in the lay-bys during the second experimental session. Each lap included three of these dummy vehicles, two in the drivers’ direction and one in the direction of the oncoming traffic.

In the following, the different critical situations and their realisation is explained in more detail.

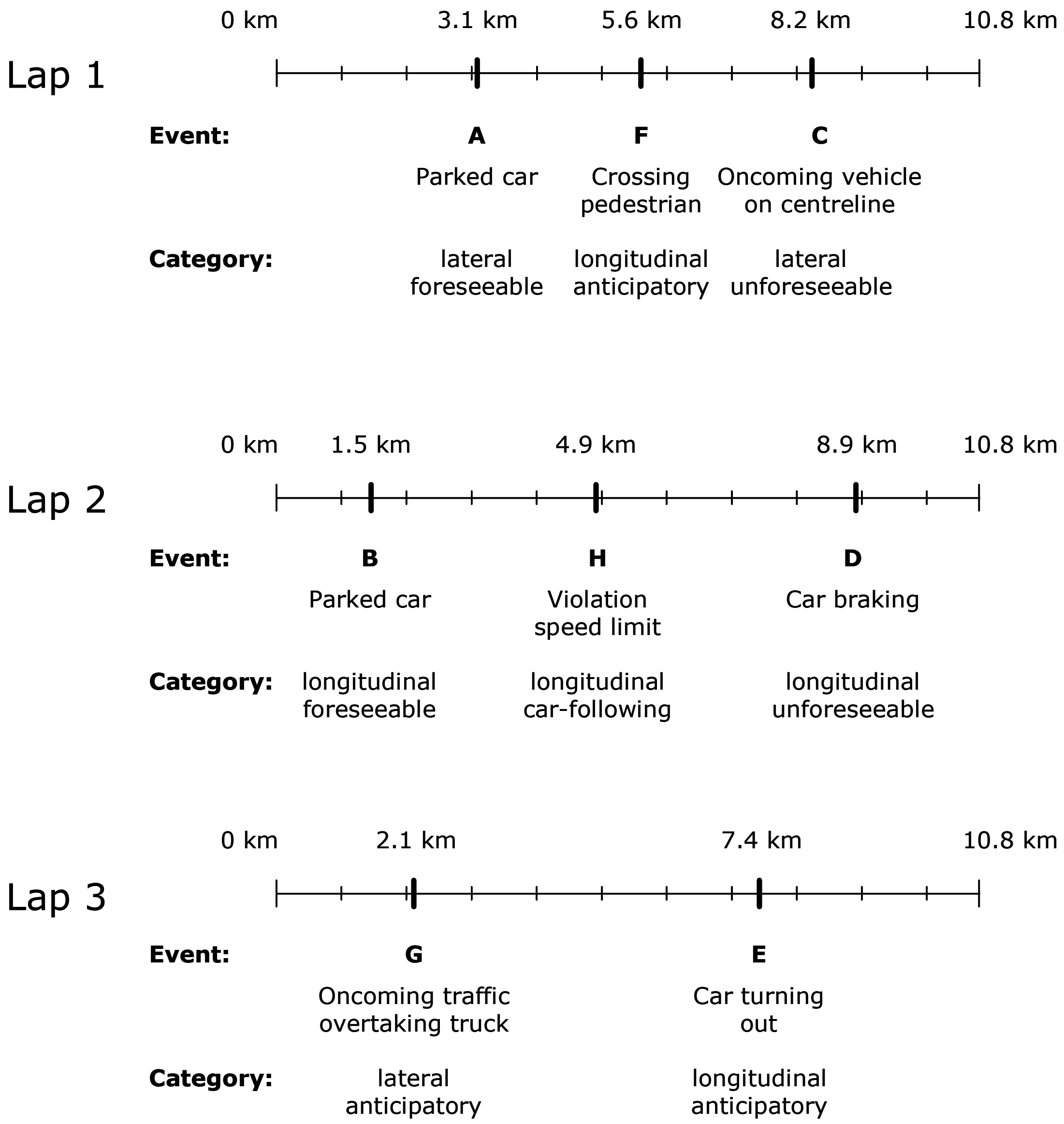


Figure 10. Approximate point of occurrence of the eight critical driving situations during the three consecutive laps of the second experimental session. The critical situations (“Events”, labelled alphabetically) belonged to different theoretically derived categories used to investigate content-specific changes in drivers’ SA as a function of the level of lane keeping assistance. The order of the three laps was counterbalanced across drivers. However, the laps always included the same Events that occurred in the same order.

3.5.1 Foreseeable situations

3.5.1.1 Lateral foreseeable - Event A

Event A was realised by a parking vehicle on the right road side partly occupying the drivers' lane (see Figure 11) with its hazard warning lamp activated. The parking vehicle was visible for drivers approximately at a distance of 300 m before. Drivers had to realise that they would have to draw aside and to overtake the parking vehicle in order to avoid a collision. Thus, this Event required adjustments in lateral control of the vehicle and was designed to simulate functional limits of the HC and LDW system. Both systems offered inadequate support in this situation because they continued in assisting drivers (to different extent) to stay within the driving lane. Thus, drivers had to recognize that they had to override the steering torques of the HC system (guiding them back towards the centre of the lane) by applying counter forces on the steering wheel, and to ignore the warning of the LDW system. Alternatively, they could switch off both systems by activating the turn indicator.

Event A was categorised as "foreseeable" because drivers had a comparable large time window to perceive and interpret the situation, and to plan ahead their response. Furthermore, the situation did not require drivers to anticipate the behaviour of other traffic participants. Instead, the situation was static and unambiguous in terms of the reaction (manoeuvre) that drivers would have to perform.

The vehicle occurred at a distance of 3,100 m in Lap 1. The vehicle was a red Volvo V70 whose left tyres were placed 2.1 m right from the road centre line.

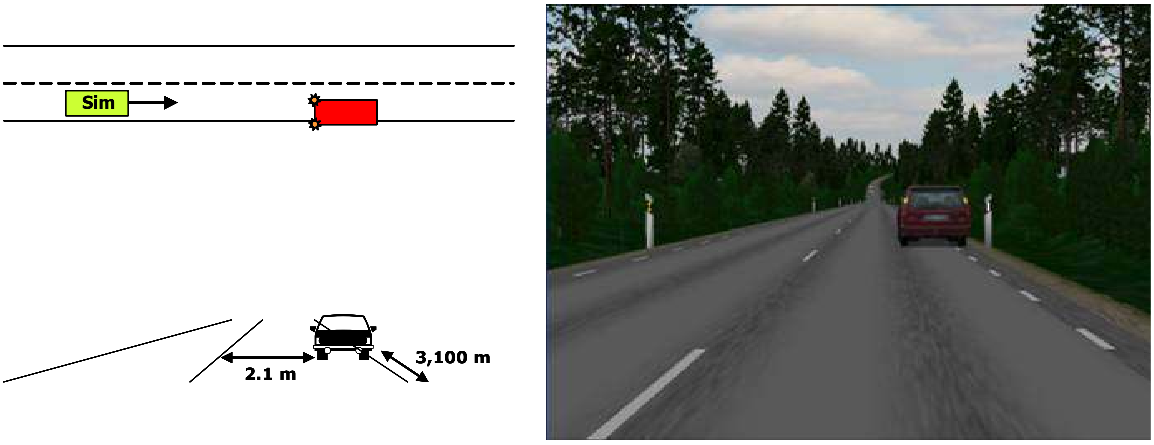


Figure 11. Pictorial view (left) and simulation screen-shot (right) of Event A from a bird's eye of view (top left) and from the driver's perspective (bottom left and right part of the figure). Drivers had to overtake the parking vehicle.

3.5.1.2 Longitudinal foreseeable – Event B

Event B occurred at a distance of 1,500 m in Lap 2. Event B was similar to Event A (a parking vehicle with its hazard warning lamp activated), but included two oncoming vehicles that were programmed in a way that drivers had to react primarily on the longitudinal control level (release of the accelerator pedal and/or braking) in order to let the vehicles pass before they could safely overtake the parking vehicle (see Figure 12). The vehicle was a white Mercedes A-class whose left tyres were placed at a distance of 2.16 m right from the road centre line. The two oncoming vehicles drove at the same speed as the driver and continued to drive at a speed of 54 km/h as soon as the drivers' speed fell below 54 km/h (e.g., when they initiated a braking manoeuvre in order to let the oncoming traffic pass).

The parking vehicle was visible for drivers approximately from 360 m before. The situation was categorised as foreseeable because it was unambiguous and demanded from drivers a clear reaction on the longitudinal control level. Furthermore drivers had a comparably large amount of time for situation perception and planning of an adequate response.

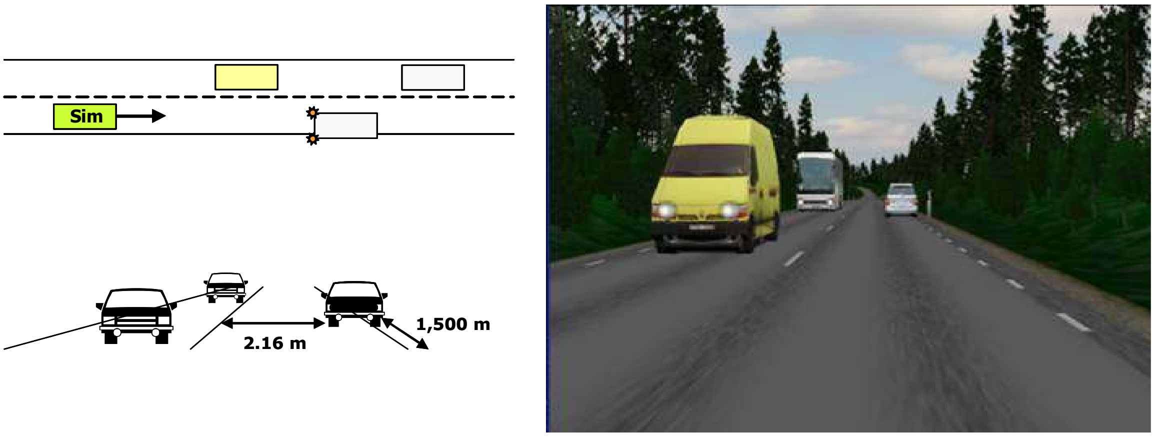


Figure 12. Pictorial view (left) and simulation screen-shot (right) of Event B from a bird's eye of view (top left) and from the driver's perspective (bottom left and right part of the figure). Drivers had to let the oncoming traffic pass before they could overtake the parking vehicle.

3.5.2 Unforeseeable situations

3.5.2.1 Lateral unforeseeable – Event C

Event C occurred in a left curve and consisted of three oncoming vehicles of which the vehicle in the middle position drove partly on the drivers' lane and had its turn indicator activated as though it intended to overtake the vehicle in front of it (see

Figure 13). Event C occurred at a distance of approximately 8,200 m (dependent on the drivers' speed) in Lap 1.

Event C was categorised as unforeseeable because the drivers' sight was obstructed due to the curve. Thus, the oncoming vehicles were first visible for drivers when the distance between drivers and the oncoming vehicles was approximately 250 m. Drivers were not able to anticipate the Event and had a rather small time window to perceive and interpret the visual cues, and for response planning and execution.

The oncoming vehicle in the middle position was a red Volvo V70 that occupied the drivers' lane to 0.2 m. The vehicles drove at the same speed as the driver. When the drivers' speed fell below 37 km/h and the distance between drivers and the red Volvo was smaller than 50 m, the speed of the oncoming vehicles was set to 72 km/h. If drivers did not decelerate to less than 37 km/h, the speed of the oncoming vehicles remained connected to the drivers' speed. In this case, the driver and the red Volvo met at a distance of 8225 m.

Event C simulated a functional limit of the HC and the LDW system. Drivers had to realise that they had to draw aside in order to avoid a potential collision with the oncoming traffic. Thus, the HC and the LDW system provided inadequate support in this situation. Drivers had to override the HC system's steering torques or to ignore the LDW system's warning in case of a nearby crossing of the outer right lane boundary. Alternatively, they could switch off both systems by activating the turn indicator in order to signal their intention to depart from the driving lane.

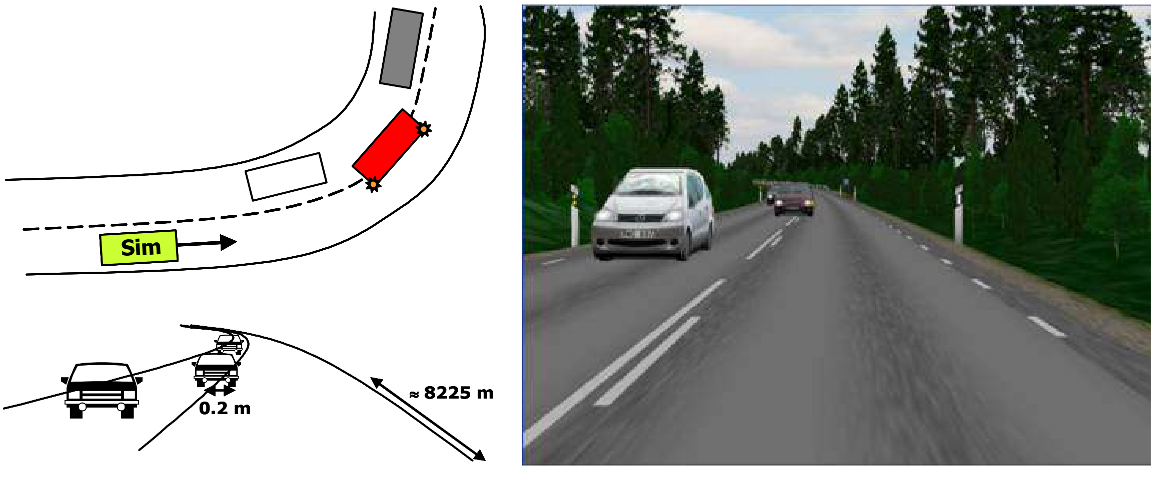


Figure 13. Pictorial view (left) and simulation screen-shot (right) of Event C from a bird's eye of view (top left) and from the driver's perspective (bottom left and right part of the figure). The oncoming vehicle in the middle position partly occupied the drivers' lane and forced drivers to draw aside in order to avoid a potential collision.

3.5.2.2 Longitudinal unforeseeable – Event D

Event D occurred at a distance of approximately 8,900 m in Lap 2. It consisted of a white Renault Master that drove in front of the driver for about 1,100 m and then suddenly set its right turn indicator, braked and turned into a lay-by at the right road side (see Figure 14).

At the beginning of the Event, the Renault parked in a lay-by at the right road side and entered the road at 7,710 m when the distance between the driver (approaching from behind) and the Renault was equal to 200 m. After that, the Renault accelerated with 2.5 m/s^2 up to a speed that was 20 km/h lower than that of the driver at the point when it had entered the road. This lower speed was chosen in order to bring the driver to drive close to the Renault in front of it so that the Renault's braking manoeuvre would occur suddenly for the driver.

The Renault stopped at 8,920 m in a lay-by with its left tyres placed 4.24 m right from the road centre line. The starting point for the activation of its right turn indicator and for the initiation of its braking manoeuvre depended on its actual speed and was set in a way that the Renault would yield the stopping point (at 8,920 m) with a constant deceleration of 2 m/s^2.

Drivers had to react primarily with respect to longitudinal control (by braking) in order to avoid a collision with the Renault. They were also free to draw aside. Event D required a sudden reaction from the driver, and thus was categorised as unforeseeable.

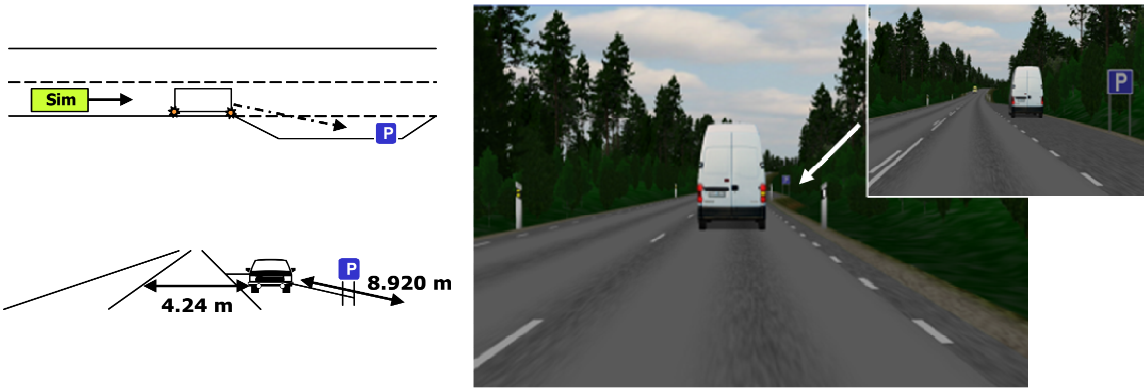


Figure 14. Pictorial view (left) and simulation screen-shot (right) of Event D from a bird's eye of view (top left) and from the driver's perspective (bottom left and right part of the figure). The vehicle in front of the driver started to brake suddenly and to turn into a lay-by at the right road side.

3.5.3 Anticipatory situations

3.5.3.1 Lateral anticipatory – Event G

Event G occurred at 2,100 m in Lap 3 and consisted of a defective truck at the left road side that partly occupied the opposing lane so that the oncoming traffic had to overtake the defective truck and while doing this, partly occupied the drivers’ lane (see Figure 15).

The defective truck was placed on the top of an ascending slope so that it was visible for drivers long (approximately 500 m) before they approached the truck. There were several oncoming vehicles that overtook the truck before drivers came close to it. Thus, if drivers perceived and interpreted the situation correctly, they were able to anticipate that they would eventually have to draw aside at the moment they passed the truck because of oncoming traffic overtaking it and partly occupying the drivers’ lane. Therefore, Event G was categorised as high in anticipation demands. It was assumed that insufficient processing of the visual cues would lead to an incomprehensive incorporation of that information in the situation model (lower SA), thus making anticipation difficult and leading to delayed and more abrupt driver reactions in that situation.

About 350 m before drivers approached the truck two oncoming vehicles overtook the truck and partly occupied the drivers’ lane. Behind them, there were two other vehicles that were programmed in a way that they overtook the truck just at the moment as the drivers approached it. These two vehicles were the ones that were “critical” for drivers because they required drivers to draw aside in order to avoid a collision. Event G simulated functional limits of the HC and LDW system because it required drivers to react on the lateral control level. Drivers had to override the HC system and to ignore the LDW system’s warnings. Alternatively they could switch off both systems by activating the turn indicator.

The first of the two critical vehicles was a white Mercedes A-class that approached the truck from the opposing direction with the same speed as the driver. When the distance between the driver and the Mercedes became less than 150 m, the speed of the Mercedes was set to 36 km/h. When the Mercedes was 60 m behind the truck, its turn indicator (left) was activated and it started to pull out behind the truck. The second critical vehicle followed the Mercedes at a distance of 20 m. While passing the truck, both critical vehicles occupied the drivers’ lane to about 0.5 m.

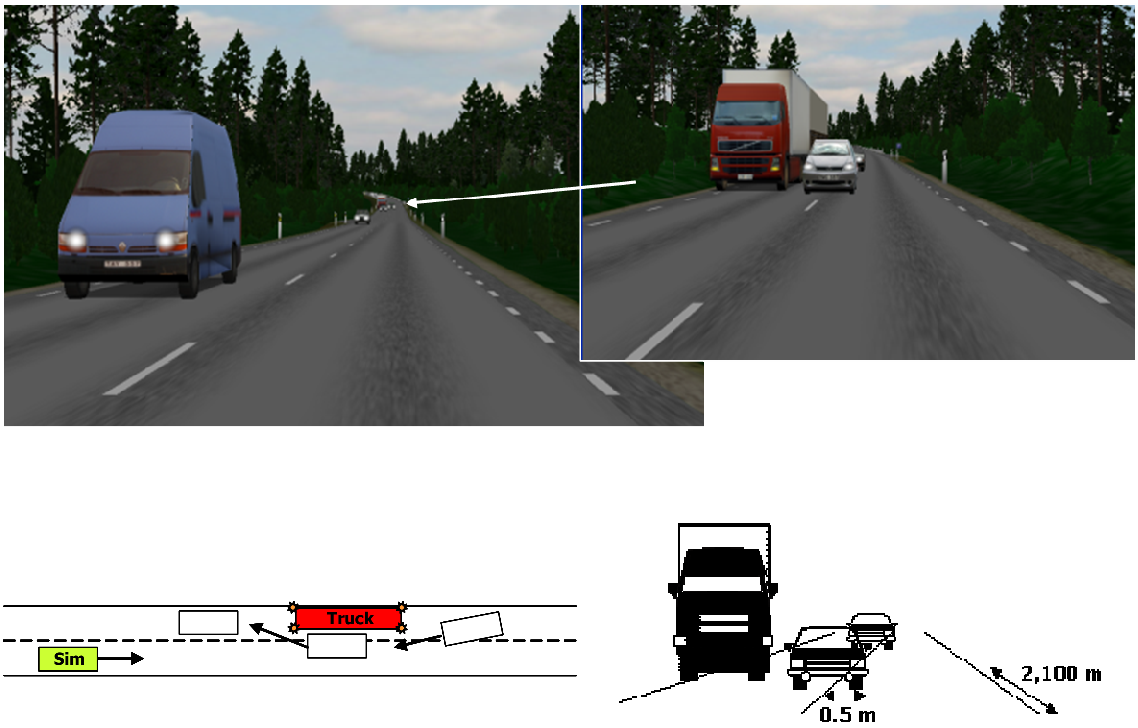


Figure 15. Top: simulation screen-shots of Event G from a driver's view about 350 m before passing the defective truck on the left roadside (top left panel) and just before approaching the truck with two oncoming overtaking vehicles partly occupying the drivers' lane (top right panel). Bottom: pictorial views of Event G from a bird's eye of view (bottom left panel) and from a driver's point of view (bottom right panel).

3.5.3.2 Longitudinal anticipatory – Event E

Event E occurred at 7,420 m in Lap 3 and consisted of a vehicle that pulled out in front of drivers' from a lay-by at the right road side (see Figure 16). The vehicle was a white Volvo V240 positioned at 7,410 m. At a Time to Collision (TTC) of 5 s the Volvo started to accelerate with 2.5 m/s^2 within the lay-by; and, after 10 m, set its turn indicator and started to move left towards the drivers' lane. The Volvo reached its final lateral position (at the centre of the lane) 44.5 m after it had started to move towards the drivers' lane. Its speed then was 59.2 km/h. The Volvo continued to accelerate constantly with 2.5 m/s^2 and finally disappeared in front of the driver.

Event E was designed in a way that should allow drivers to anticipate the intention of the Volvo to pull out in front of them based on its acceleration in the lay-by. The Volvo drove straight ahead for 2.7 s in the lay-by before it started to move towards the drivers' lane. It was assumed that an insufficient processing of the visual cues would affect the comprehensiveness and accuracy of drivers' mental representation of the situation, which in turn should impair drivers' anticipation of the future development of the situation. Thus, lower levels of SA should become apparent in later and more abrupt driver reactions in that situation because of an incomplete incorporation of the relevant situation elements in the situation model.

Event E required drivers to react primarily on the longitudinal control level. However, drivers could principally also draw aside in order to avoid a collision.

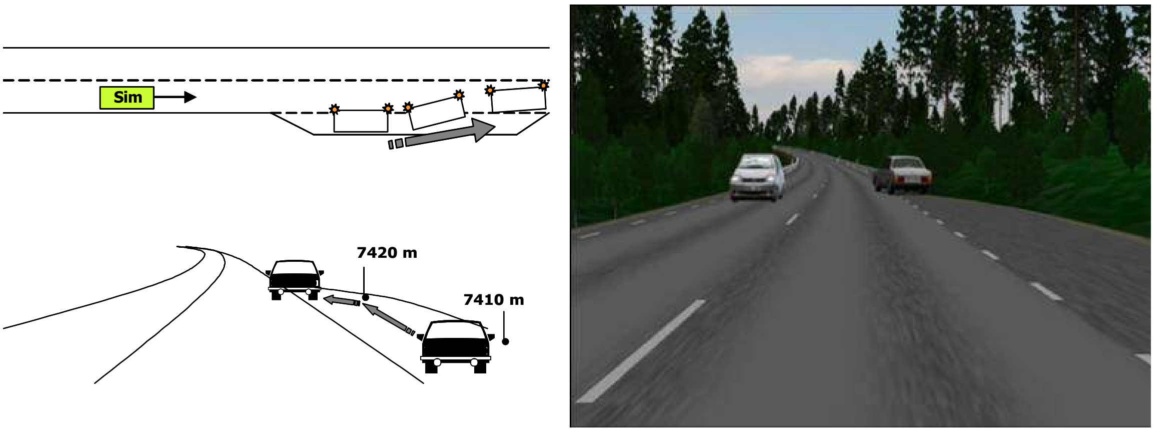


Figure 16. Pictorial view (left) and simulation screen-shot (right) of Event E from a bird's eye of view (top left) and from the driver's perspective (bottom left and right part of the figure). The vehicle starts to accelerate and to drive straight ahead in the lay-by for 2.7 s before moving towards the drivers' lane and entering the road in front of the drivers.

3.5.3.3 Longitudinal anticipatory – Event F

Event F occurred at 5,600 m in Lap 1 and consisted of a pedestrian crossing the road in front of drivers. The starting position of the pedestrian was 6.5 m right from the road "in the woods". On the left side of the road there was a bus parking in a lay-by with its left tyres placed 3.95 m left from the centreline. A group of four people was standing in front of the bus (see Figure 17). At a Time to Collision (TTC) of 6 s with the pedestrian, the pedestrian started to walk towards the group of people with a speed of 1.67 m/s. There was no oncoming traffic in this situation.

The pedestrian walked a distance of 6.5 m up from its starting position at the right side of the road before it entered the road. Based on its movement towards the road as well as the placement of the bus and the people at the left road side, drivers should be able to anticipate the future development of the situation and to infer the intention of the pedestrian to cross the road. Thus, the situation was categorised as requiring a high amount of anticipation. It was assumed that an insufficient processing of the visual cues should lead to an insufficient understanding of the situation (impaired SA) which in turn should hinder an accurate anticipation of the situation. This should in turn become apparent in drivers' delayed and more abrupt reaction to that situation.

Drivers were assumed to react primarily with respect to longitudinal control (braking). However, they could also draw aside in order to avoid a potential collision with the pedestrian.

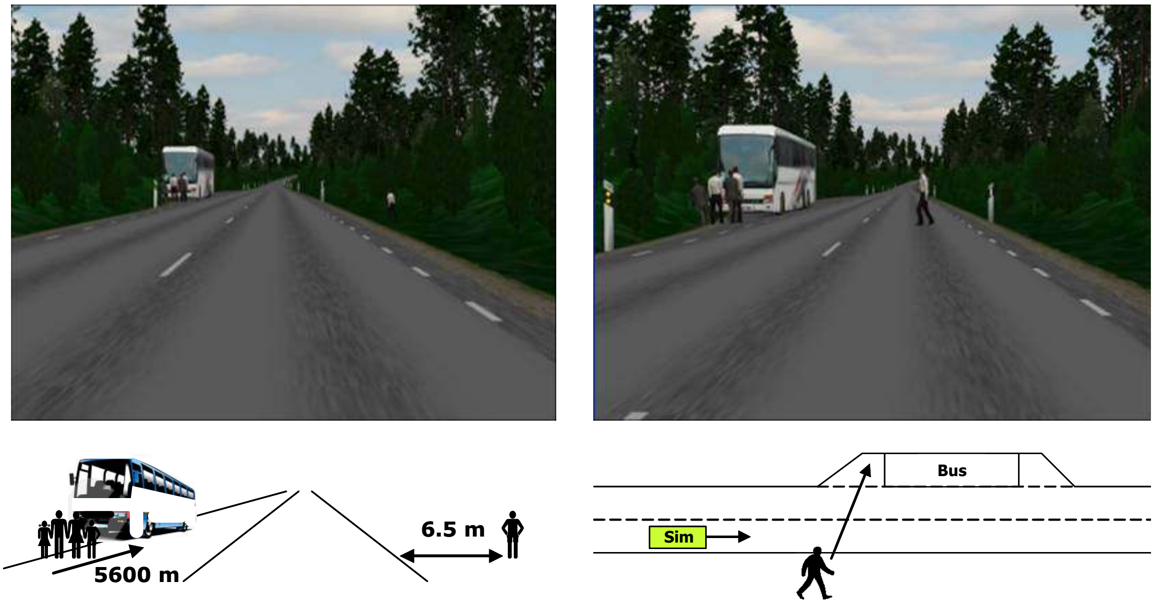


Figure 17. Top: simulation screen-shots of Event F. Top left panel: drivers' view when they approached the Event. The pedestrian started to walk at TTC = 6 s towards the road from a position of 6.5 m right from the road. Top right panel: drivers' view when they further approached the Event. The pedestrian crosses the road and walks towards the group of people positioned in front of the bus. Bottom: pictorial view of Event F from a driver's point of view (bottom left) and from a bird's eye of view (bottom right).

3.5.4 "Car-following" situation – Event H

Event H was the only situation that was categorised as "not safety-critical" in terms of demanding a reaction from drivers in order to avoid a potential collision. Event H instead was designed to assess drivers' SA on the basis of their tendency to follow the erroneous behaviour of a lead vehicle which broke a speed limit.

The speed limit sign (50 km/h) was positioned at 4,900 m in Lap 2. Event H started with a vehicle (a yellow Renault Master) that approached drivers from behind with a higher speed; then overtook drivers and drove for about 1 km in front of them with the same speed as the drivers in order to keep a constant distance between the two vehicles. When the Renault passed the 50 km/h speed limit sign, it did not reduce its speed accordingly, but continued to drive with a constant speed that it had before passing the sign. At 5,500 m, there was another sign that set back the speed limit to 70 km/h. After passage of this sign, the Renault accelerated with 4 m/s^2 and finally disappeared in front of drivers. Event H is depicted in Figure 18.

It was assumed that drivers' reactions on the longitudinal control level (speed reduction) at passage of the 50 km/h speed sign would indicate their correct perception and interpretation of the situation, and thus, indicate a higher level of SA. Event H was the only situation in which drivers' SA was *not* assessed by the timeliness and the abruptness of their reaction, but by their tendency to show a reaction in response to the changing speed limit or not.

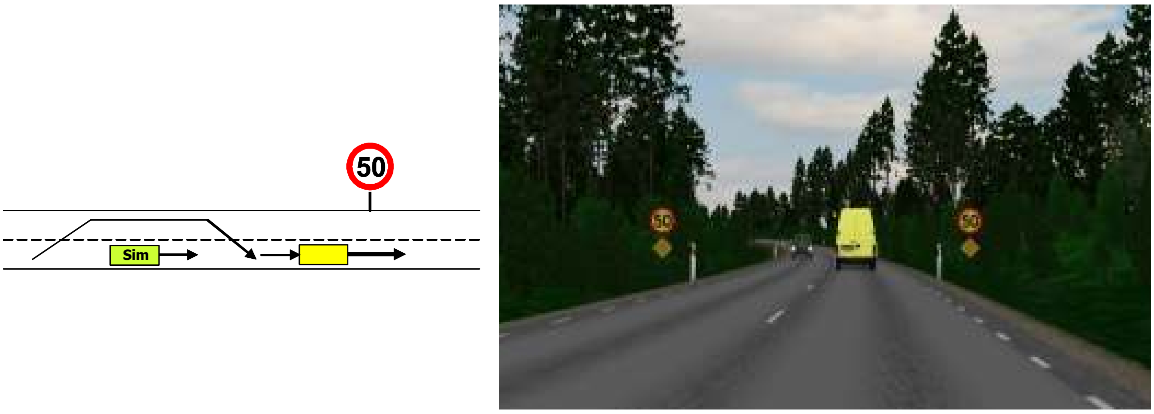


Figure 18. Pictogram from a bird's eye view (left) and simulation screen-shot (right) of Event H. The yellow vehicle in front of the drivers violates the 50 km/h speed limit and continues to drive with a higher speed.

3.6 Experimental Design

A mixed two-factor repeated measures design was used in this study. The *first factor "level of lane keeping assistance"* had three levels and was varied between-subjects. Drivers were assigned to one of three experimental groups (yielding 15 drivers in each group) corresponding to different levels of lane keeping assistance. Drivers who drove with a high level of lane keeping assistance were assisted by a HC system, whereas drivers who drove with a low level of lane keeping assistance were assisted by a LDW system. Drivers in the control group drove without lane keeping assistance system (No Assistance, NA).

The *second factor "session"* had two levels and was varied within-subjects. The session factor served as manipulation of drivers' experience of functional system limits. Critical driving situations (including situations that simulated functional system limits) occurred only during the second experimental session. In Session 1 however, drivers encountered no critical driving situations and thus, did not experience functional system limits.

Measures of drivers' reliance on the lane keeping assistance systems and of the hypothesised mediating factors (drivers' attitudes, energetic and motivational processes) were observed both as a function of the level of lane keeping assistance (over the levels of the between-subjects factor) *and* as a function the occurrence vs. non-occurrence of functional system limits (over the levels of the within-subjects factor).[3]

The second major concept of interest in this study – drivers' SA – was solely investigated as a function of the level of lane keeping assistance. Drivers' SA was assessed by their reaction to certain critical driving situations, and thus, was only measured during the second experimental session. Accordingly, for the investigation of drivers' SA, a one-factor between-subjects design (with the three levels of lane keeping assistance) was used.

In order to verify the beneficial effect of the level of lane keeping assistance on drivers' lane keeping performance (manipulation check), a 3 (level of lane keeping assistance) x 2 (session) x 2 (arrows task; driving with vs. driving without concurrent performance of the Arrows Task) mixed factorial design was used, in which the two factors "session" and "arrows task" were varied within-subjects.

3.7 Dependent Measures

The first major aim of this study was to investigate changes in drivers' reliance on a lane keeping assistance system as a function of the level of assistance (level of automation of the lateral control driving subtask) and as a function of drivers' experience of functional system limits. It was hypothesised that reliance on automation is influenced by three major factors, which are drivers' attitudes towards the assistance systems (trust), as well as energetic (arousal) and motivational processes (mental effort regulation). In order to study these assumed relationships, a range of objective (behavioural) and subjective (questionnaire) measures were collected.

The second major aim of the study was to investigate changes in drivers' SA as a function of the level of automation of the lateral control driving task (level of lane keeping assistance). SA was assessed by objective performance-based measures of drivers' reactions to critical driving situations, as well as by subjective measures.

The questionnaires given to subjects consisted of existing, validated scales as well as of newly developed scales and items specifically designed for this purpose of research. For the existing scales, Swedish versions already used in previous driving studies were used for this experiment. All other scales and items were first developed

---

[3] In order to investigate the effects of the level of lane keeping assistance and of drivers' experience of functional system limits on reliance and mediating variables, both objective (performance) and subjective measures were collected. The performance measures (of drivers' reliance and of mental effort regulation) were analysed over *three* levels of the within-factor (Session 1, Session 2 during normal driving periods, Session 2 during critical driving situations). Subjective measures were collected after the end of both experimental sessions, and thus, no differentiation could be made for these measures between periods of occurrence vs. non-occurrence of critical driving situations in Session 2.

in English, and then translated into Swedish by one researcher at VTI. Afterwards, the Swedish translations were cross-checked by another independent Swedish researcher, and translated back into English by a third researcher in order to verify the correct understanding of the items.

3.7.1 Drivers' lane keeping performance

The evaluation of drivers' lane keeping performance served primarily as manipulation check of the beneficial effect of increasing levels of lane keeping assistance on drivers' lane keeping performance. The following two performance measures were evaluated: the mean standard deviation of lateral position and the mean duration of the HC system's maximum steering torque application.[4]

3.7.2 Drivers' reliance on the lane keeping assistance systems

Process-oriented performance-based measures were used in order to study drivers' reliance on the lane keeping assistance system. In particular, drivers' strategies for allocating their visual attention between the road scene (corresponding to the most relevant source of information while driving) and the visual display of the secondary task (Arrows Task, see chapter 3.4) served as a measure for their reliance on the lane keeping assistance systems. It was assumed that drivers' increasing reliance on the lane keeping assistance systems would be reflected by their increased preparedness to allocate their visual attention away from the road scene to the visual display of the Arrows Task. Two measures of drivers' visual attention allocation strategies were assumed to be sensitive to different levels of reliance, namely the *mean duration of drivers' single glances to the Arrows Task display*, and the *mean proportion of display glance time* (proportion of time that drivers glanced to the in-vehicle visual display until they responded to the single displays of the Arrows Task). It was expected that drivers who rely to a high extent on a lane keeping assistance system take the risk of longer glances away from the road scene to the Arrows Task's visual display (resulting in an increased mean duration of single glances to the display compared to drivers in the "No Assistance" group) and devote fewer visual attentional resources to the road scene during performance of the Arrows Task (resulting in a higher proportion of display glance time).

---

[4] Although only drivers driving with the HC system actually experienced the steering torques of the HC system, the forces that the HC system would have generated if drivers were assisted by it were also recorded for LDW drivers and drivers in the control group.

Drivers' eye movements were recorded at 60 Hz using a Smart Eye Pro eye tracker system (Version 3.1) which consisted of three Sony XC HR50 monochrome cameras, equipped with two IR-illuminators. Figure 19 shows the placement of the three cameras and the image of the Smart Eye program during data recording.

Figure 19. Left panel: Position of the three cameras and the two IR lights used for tracking the drivers' eye and head movements and their gaze direction. Right panel: Image of the Smart Eye program during gaze tracking.

3.7.3 The influence of motivational processes on drivers' reliance: Mental effort regulation

Motivational processes, i.e. drivers' strategic decision about the investment of mental effort were assumed to play a major role in influencing their reliance on the lane keeping assistance systems. Previous research suggests that reliance on automation can be viewed as an effective operator strategy to limit the expenditure of mental effort while maintaining an acceptable task performance due to the support offered by the automation. Thus, it was hypothesised that high reliance on a lane keeping assistance system is influenced by drivers' attempt to conserve mental effort, which in turn should be associated with lower costs in the form of latent performance decrements. Two forms of latent performance decrements were supposed to play a role in this study: compensatory costs and subsidiary task failures (Hockey, 1997). In previous driving studies, measures of lateral vehicle control (heading error, steering reversals) have been shown sensitive for drivers' strategic regulation of mental effort (Fairclough, 2001; Matthews & Desmond, 2002). However, lateral control measures were confounded by the effects of the lane keeping assistance systems in this study and could not be used as clear indicators for drivers' strategic regulation of mental effort. For this reason, several questionnaire items were developed in order to identify the mental effort regulation strategies that drivers may have adopted when driving with different levels of lane keeping assistance.

3.7.3.1 Compensatory costs

Overall mental workload

Drivers' overall mental workload was measured with the Rating Scale Mental Effort (RSME; Zijlstra, 1993) which is a one-dimensional scale with ratings between 0 and 150 and nine descriptive indicators along its axis (ranging from "absolutely no effort" to "extreme effort" - see page 3 in Appendix A). The RSME has been used in a number of driving studies and has proved sensitive for variations in task demands (e.g., De Waard, 1996; Verwey & Veltman, 1996).

As a second measure of overall mental workload, the Raw Task Load Index (RTLX; Byers, Bittner, & Hill, 1989) was used. The RTLX constitutes the unweighted mean of drivers' workload ratings on the six dimensions of the NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988). The six dimensions of workload are assessed on continuous bipolar rating scales ranging from 0 to 100 (see page 6-7 in Appendix A). The NASA-TLX has the additional advantage that the primary contributors to the overall workload index can be estimated by referring to the different workload dimensions.

Frustration

Drivers' level of frustration was measured by means of the respective subscale of the NASA-TLX.

3.7.3.2 Subsidiary task failures

Subsidiary task failures – the selective impairment of peripheral task components – were assessed by evaluating drivers' performance in the secondary task. Measures of drivers' performance in the Arrows Task were response accuracy (percentage of correct, false, and missing responses) and response time.

3.7.3.3 Mental effort regulation strategies

Several questionnaire items were developed in order to find out the strategies that drivers had adopted in the regulation of their mental effort.

One item addressed drivers' relative prioritisation of the driving task and the secondary task by asking them to mark on a continuous line (corresponding to their overall mental effort) how much effort they had proportionally invested in the driving task vs. in the Arrows Task during the preceding drive (see page 4 in Appendix A).

The other items had the form of 7-point rating scales with the opposite poles "strongly agree" and "strongly disagree". Table 5 shows three example items.

Table 5. Examples of items used to assess drivers' mental effort regulation strategies.

	Strongly disagree						*Strongly agree*
I tried to maintain an acceptable level of driving performance without investing too much effort in it.	☐	☐	☐	☐	☐	☐	☐
It was important for me that the Arrows Task did not have a negative effect on my driving performance.	☐	☐	☐	☐	☐	☐	☐
I could not perform as well as I wanted on the Arrows Task. The driving task required all of my attention.	☐	☐	☐	☐	☐	☐	☐

3.7.4 The influence of drivers' attitudes on their reliance: Drivers' trust in the lane keeping assistance systems

Drivers' trust in the HC and in the LDW system was measured on a bipolar continuous rating scale ("How much did you trust the system?") used by Muir and Moray (1996) ranging from "not at all" to "extremely high". Furthermore, drivers were asked to rate different aspects of the systems' behaviour that are assumed to constitute the basic information upon which trust is built (Lee & See, 2004; Muir & Moray, 1996). Performance-related information was evaluated by asking drivers to rate the systems' *competence* and *predictability*, whereas process-related information was evaluated by asking drivers to rate the *comprehensibility* of the systems' behaviour. Subsequent analysis were planned to determine the relative importance of these three types of system-related information for the development of drivers' trust in the lane keeping assistance systems. The scales are presented on page 10 of Appendix A.

3.7.5 The influence of energetic processes on drivers' reliance: Arousal

Drivers' arousal was measured with the Stanford Sleepiness Scale (SSS; Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973) which is a 7-point rating scale with descriptors on each point of the scale. Higher scores on this scale indicate a lower level of subjective arousal (see Figure 20).

1	Feeling active and vital. Alert and wide awake.
2	Functioning at a high level, but not at peak. Able to concentrate.
3	Relaxed and awake, but not at full alertness. Responsive.
4	A little foggy, not at peak. Let down.
5	More foggy. Beginning to lose interest in staying awake. Slowed down.
6	Very sleepy, fighting sleep, woozy. Prefer to be lying down.
7	Almost asleep. Lost struggle to remain awake.

Figure 20. Stanford Sleepiness Scale (Hoddes et al., 1973).

3.7.6 Evaluation of system usability and safety effects

In order to yield a more complete picture about drivers' adaptation processes in response to the automation of the lateral control task, a major part of the questionnaire dealt with drivers' evaluation of the lane keeping assistance systems. For this purpose, drivers' were asked to evaluate the HC system and the LDW system in terms of the adequacy of their interventions in driving, their usability and controllability, and in terms of their supposed effects on driving behaviour and safety. Items concerning system usability were adapted from the Questionnaire for User Interaction Satisfaction (QUIS; Chin, Diehl, & Norman, 1988; Shneiderman & Plaisant, 2005). Items for assessing controllability were based on descriptions in the RESPONSE Code of Practice for the Design and Evaluation of ADAS ("RESPONSE 3 Code of practice for the design and evaluation of ADAS," 2006). Most of these items had the form of 7-point rating scales. Examples are presented in Table 6.

Table 6. Example questionnaire items used for the evaluation of the lane keeping assistance systems in terms of usability and safety.

Adequacy of the systems' interventions in driving	The system intervened ☐ too early ☐ in due time ☐ too late
	The system's actions were ☐ too strong ☐ appropriate ☐ too weak
Usability	I needed a lot of time to learn how the system worked. (1 - strongly disagree; 7 - strongly agree)
	It was easy to drive with this system. (1 - strongly disagree; 7 - strongly agree)
Controllability	Generally, I felt in control of things. (1 - strongly disagree; 7 - strongly agree)
	I had no problems to take over control over the system when it was necessary. (1 - strongly disagree; 7 - strongly agree)
Effects on driving behaviour and safety	I think the system would encourage me to pay more attention to things unrelated to driving. (1 - strongly disagree; 7 - strongly agree)
	I think the system would encourage me to drive riskier. (1 - strongly disagree; 7 - strongly agree)
	I think driving with this system would make driving more monotonous. (1 - strongly disagree; 7 - strongly agree)

3.7.7 Individual driver characteristics

Before participants started to drive in the simulator, they were first asked to fill in a demographic questionnaire with questions about their driving experience, their experience with and their attitude towards driver assistance systems, their interest in technology, and their driving style. Those driver variables were assumed to have potentially an influence on drivers' adaptation processes in response to different levels of automation of the lateral control task. The questionnaire can be found in Appendix B.

3.7.8 Situation awareness

Performance-based measures of drivers' reactions to different types of critical driving situations that occurred unexpectedly during the second experimental session served as indicators for drivers' SA. More specifically, the relative point in time of drivers' initial reaction to the situations and the abruptness of their initiated driving manoeuvre served as the major dependent variables in order to assess drivers' SA. Thus, two types of SA measures were analysed: (a) Measures that indicated the relative point in time of drivers' initial reaction to the critical driving situations and (b) measures that indicated the abruptness of the driving manoeuvre that was initiated by drivers in response to their encountering of the situations.

In order to yield a more general picture of the effects of the level of lane keeping assistance on drivers' behaviour in safety-critical driving situations, some more general measures of driving performance were assessed for the purpose of explorative data analysis. Those measures however were not taken as indicators for drivers' SA.

Table 7 lists the standard measures that were individually analysed for each driver for each of the safety-critical situations[5].

The different nature of the critical situations required additionally the analyses of some other more situation-specific driver performance measures (those will be dealt with in the results section).

Additionally to these performance-based measures of drivers' SA, subjective measures of SA were collected by asking drivers to rate their awareness in relation to different driving task-related information on eight 7-point rating scales with the poles "not aware at all" and "completely aware" (see page 8 in Appendix A).

---

[5] In total, drivers encountered eight critical situations during the second experimental session, of which seven were „safety-critical" (requiring a response in order to avoid collisions with other traffic) and one was „not safety-critical". This latter situation (Event H, see chapter 3.5.4) assessed drivers' SA by their tendency to follow the erroneous behaviour of a lead vehicle that broke a speed limit. The SA measure in this situation was drivers' reduction of speed (yes/no) when passing the speed limit, and *not* the timeliness of drivers' reaction or the magnitude of reaction (speed change).

Table 7. Standard driving performance measures of the timeliness and abruptness of drivers’ reaction to the critical driving situations used as indicators for drivers' SA when driving with different levels of lane keeping assistance. Measures that describe the general characteristics of the initiated driving manoeuvre (last row) were analysed for explorative purposes in order to investigate potential effects of the lane keeping assistance systems on drivers’ behaviour in critical driving situations. All measures were extracted individually for each driver for each of the critical driving situations.

Variable	Operational definition	Control level
Timeliness of drivers’ reaction to critical situation	Time to Collision (TTC)[a] [s] at the initial point of drivers’ reaction on either the longitudinal control level (release of accelerator pedal, braking) or on the lateral control level (drawing aside)	lateral & longitudinal
Abruptness of initiated driving manoeuvre	Time [s] between release of accelerator pedal and application of brake	longitudinal
	Maximum braking force [N]	longitudinal
Manoeuvre characteristics	Maximum lateral deviation [m] / Magnitude of change in lateral position [m]	lateral
	Minimum speed [km/h] / Magnitude of speed change [km/h]	longitudinal

[a]TTC was calculated as the time to drivers’ collision with situation-specific stationary or moving objects, e.g. oncoming vehicles, or vehicles occupying the drivers’ lane. In the case of stationary objects, Time Headway (THW) was used instead of TTC.

3.8 Procedure

Drivers were recruited by telephone by the experimental leader based on predefined participation criteria (a minimum total driven distance of 10,000 km, no spectacle wearers etc.) from a pool of Swedish drivers that had expressed their interest in taking part in driving simulator studies at VTI. Participants were matched for gender and as much as possible for age and than assigned to one of the three experimental conditions (driving with the HC system corresponding to a high level of lane keeping assistance, driving with the LDW system corresponding to a low level of lane keeping assistance, driving without lane keeping assistance).

When participants arrived at VTI in Linköping they were welcomed by the experimental leader and first given a written instruction (see Appendix C - E) that informed them about the aim and the procedure of the study, their tasks during the study, and the functionality of the lane keeping assistance systems (except for drivers in the control group). Participants were told that the aim of this study was to investigate “drivers’ situation awareness on rural roads under varying conditions”. They were informed about the speed limit and instructed to drive as they normally would on a road with corresponding character concerning interaction with other traffic participants, manoeuvring, and time gaps. Participants in the HC and LDW group were informed about the functionality of the lane keeping assistance systems and instructed to use the turn indicator when they had the intention to depart from the driving lane, e.g. when they wanted to perform a lane change manoeuvre. They were told that the activation of the turn indicator would suppress the warnings of the LDW system and deactivate the HC system’s steering torques. Participants in the HC group were additionally told that they could principally override the system if they applied counter forces to the steering wheel greater than the maximum steering force of the HC system. They were also informed that the HC system would be automatically deactivated when the vehicle was about half a meter outside the lane boundaries. Participants in the HC condition were instructed to keep their hands on the steering wheel in order to be able to experience the support from the HC system. They were also informed that the HC system’s steering torques would not be in any case strong enough to guide the vehicle completely back to the centre of the lane without their help.

Participants were then informed that they would have to perform a second task while driving: the Arrows Task. They were instructed to perform the Arrows Task as well and as fast as possible, but without jeopardising traffic safety. It was emphasised that the participants’ primary responsibility was to drive the car in a normal and safe way, and that they would have to decide on their own how much attention they could allocate to the secondary task while driving.

After participants were made familiar with the study, they were informed that various data about their performance would be collected. Participants were assured that this data would only be used for scientific purposes and would be treated strictly confidential, making it impossible to identify how individual persons had performed, behaved or answered.

Participants were informed that the experimental leader would see and hear them when they drove in the simulator vehicle, and that they could at any time communicate with him or her via the microphone installed in the car (see Figure 21). Participants were asked to inform the experimental leader as soon as they experienced any kind of discomfort during the drive. They were also informed that they could at any time interrupt their participation in the study without having to give any explanation.

After it was made sure that participants understood all the information about the study and got all their questions answered, they signed an informed consent form confirming their willingness to participate in the study.

Figure 21. Desk of the experimental leader during the experiment. The experimental leader did supervise the proper progress of the study and the participants’ behaviour via multiple monitors. Participants could at any time talk to the experimental leader via a microphone.

After having filled-in the informed consent, participants answered a demographic questionnaire (see Appendix B). Then they took place in the simulator where the experimental leader explained them the car and how to drive with the simulator. The experimental leader shortly repeated again the procedure of the study and told participants that they were free to ask any questions during the following practice session.

Before the practice session started, a personal profile of each drivers’ head and face was created with the Smart Eye pro eye tracking system in order to enable the recording of the participants’ eye movements during the drives.

After that, participants were again reminded of the speed limit and started the first practice session of 10.8 km length. The purpose of this practise session was to get participants acquainted with driving in the simulator and with the behaviour of the lane keeping assistance systems. The experimental leader reminded participants again of the functionality of the HC and LDW system and invited them to test the systems’ actions by driving closer to the lane boundaries and performing some lane change manoeuvres with and without activating the turn indicator. This should enable drivers to build up a sufficient understanding of the lane keeping assistance systems (a precondition for the development of trust in the system). Participants in the control group were also invited to perform some lane change manoeuvres in order to get used to the feeling of driving in the simulator.

At the end of the training session participants were asked to fill-in a second short questionnaire and to rate their level of arousal and their overall mental effort on the SSS and RSME, respectively.

After that, participants were made familiar with the secondary task while the simulator vehicle was parked. The experimental leader asked participants to perform a number of practice trials (10 trials minimum) until they had the feeling that they had understood the task. Participants were also instructed to give intentionally the wrong answer in order to experience the auditory performance feedback.

After practicing the secondary task, participants started with the first experimental session which consisted of two laps of the experimental route yielding a total length of 21.6 km (about 15 min driving time). During this session, participants drove either with the HC system, the LDW system, or without lane keeping assistance system (control group) and concurrently performed the Arrows Task while driving. During the drive, participants' eye movements, their secondary task performance, and their driving performance were recorded.

After the first experimental session the simulation was stopped and participants were asked to fill-in a third questionnaire including all scales and items except those referring to the evaluation of system usability, acceptance, and safety effects[6].

Then the second experimental session started. During this session participants drove the experimental route for three times corresponding to an overall length of 32.4 km (about 20-25 min driving time). As in the first experimental session, participants drove with one of three levels of lane keeping assistance and concurrently performed the Arrows Task while driving. Furthermore they unexpectedly encountered eight critical driving situations. Those situations were designed to partly simulate functional system limits in order to investigate changes in drivers' reliance on the lane keeping assistance systems as a response to their experience of these system limits. In addition, these critical situations were designed to investigate changes in drivers' SA as a function of the level of automation of the lateral control task (level of lane keeping assistance). During each lap, drivers encountered two up to three of these critical situations which always occurred in the same order. The order of the three laps was counterbalanced across participants. In Session 2, the same performance measures were collected as in the first experimental session. Additionally, performance measures of drivers' reactions to the critical driving situations were recorded as indicators for their SA.

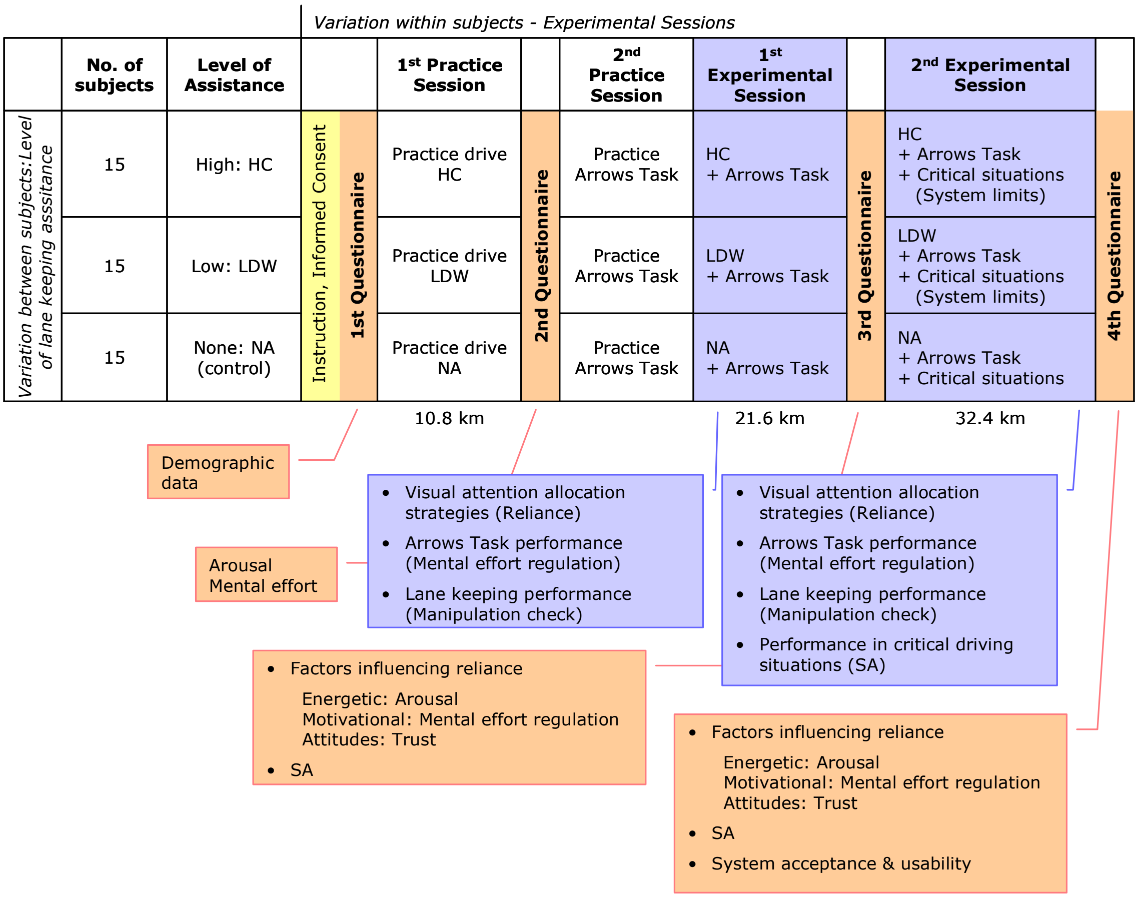
After the second experimental session participants quit the simulator and filled-in the fourth questionnaire. The fourth questionnaire consisted of the same questions and scales as the third questionnaire, but additionally included items concerning system acceptance, usability, participants' feeling of realism in the driving simulator and their evaluation of possible effects of the lane keeping assistance systems on long-term driving behaviour and safety.

After that the experiment ended. Participants were debriefed and received 400 SEK for their participation.

Figure 22 summarises the experimental design as well as the main dependent variables and measures that were collected during the course of the study.

---

[6] System-related items were left out from the questionnaires for drivers in the control group.



	No. of subjects	**Level of Assistance**			**1st Practice Session**		**2nd Practice Session**	**1st Experimental Session**		**2nd Experimental Session**	
Variation between subjects:Level of lane keeping assitance	15	High: HC	Instruction, Informed Consent	**1st Questionnaire**	Practice drive HC	**2nd Questionnaire**	Practice Arrows Task	HC + Arrows Task	**3rd Questionnaire**	HC + Arrows Task + Critical situations (System limits)	**4th Questionnaire**
	15	Low: LDW			Practice drive LDW		Practice Arrows Task	LDW + Arrows Task		LDW + Arrows Task + Critical situations (System limits)	
	15	None: NA (control)			Practice drive NA		Practice Arrows Task	NA + Arrows Task		NA + Arrows Task + Critical situations	

Figure 22. Experimental design of the study. The frames below list the main dependent measures that were collected.

4 RESULTS

This chapter is organised in two large parts. Part one deals with results on drivers' reliance on the lane keeping assistance systems including mediating factors (drivers' trust in the systems, mental effort regulation, and energetic arousal). Part two deals with drivers' reaction to the critical driving situations in Session 2 that was used as indicator for their SA.

4.1 Statistical Analyses

Planned contrasts (Furr & Rosenthal, 2003; Rosenthal, Rosnow, & Rubin, 2000) were conducted in order to evaluate the degree to which the data observed for the various measures followed the predicted effects of the level of lane keeping assistance (manipulated between-subjects) and of drivers' experience of functional system limits (manipulated within-subjects). Contrast analysis allows for analysing specific, theoretical derived predictions with a higher power compared to unfocused, omnibus tests (e.g., analyses of variance) and for obtaining an effect size that directly addresses a given research hypothesis.

As effect size measure for contrasts r is reported. Correlational effect sizes can be interpreted as the degree to which the obtained pattern of data matches the theoretically-expected pattern of data. In the case of contrasts involving the repeated factor, r was derived from *Hedges' g* corrected for the correlation between levels of the repeated factor (Dunlap, Cortina, Vaslow, & Burke, 1996; Sedlmeier & Renkewitz, 2008). Thus, r can be directly compared for contrasts involving the between-subjects factor and the within-subjects factor.

In the case of unpredicted findings, unfocused ANOVA tests followed by post-hoc comparisons (with Bonferroni-corrected alpha levels) were used as tests for statistical significance. As effect size measure for tests with more than one numerator degree of freedom, generalised eta squared (η^2_G) is reported which provides comparability across between-subjects and within-subjects designs (Bakeman, 2005; Olejnik & Algina, 2003).

An alpha level of .05 was used for all statistical tests.

4.2 Drivers' Lane keeping Performance

The evaluation of drivers' lane keeping performance served primarily as manipulation check. Drivers' lane keeping performance was observed in periods with and without concurrent secondary task performance over both experimental sessions. Thus, a 3 (level of lane keeping assistance: HC, LDW, NA) x 2 (session) x 2 (arrows task: driving with vs. without concurrent performance of the Arrows Task) mixed repeated measures design was used for the analysis of drivers' lane keeping performance; whereby the level of lane keeping assistance was varied between-subjects and the two factors session and arrows task were varied within-subjects.

Measures of drivers' lane keeping performance were extracted from at least 10 s-long periods of driving with and without concurrent Arrows Task performance (Östlund et al., 2005). In Session 2, only Arrows Tasks were included where *no* critical driving situations occurred. Furthermore, road segments around 200 m before and after the six "dummy" vehicles positioned in lay-bys at the right road sight were excluded from the analysis. Those "dummy" vehicles were introduced in Session 2 in order to avoid that drivers could infer that "something will happen" each time when there were other vehicles standing in the lay-bys[7]. Some drivers were found to react to these vehicles by changes in lateral control (e.g., by keeping a large lateral distance to them).

In order to obtain one single value for each driver for each combination of the two within-subjects factors, lane keeping performance measures were aggregated over around 19 periods of driving for each of the four combinations of the two within-subjects factors (Session 1 without and with Arrows Task, Session 2 without and with Arrows Task). The driving periods had an average duration of 28.66 s (*SD* = 5.71) and a mean length of 595.16 m (*SD* = 129.08).

It was assumed that drivers' lane keeping performance improves as a function of the level of lane keeping assistance. Thus, a main effect of the level of lane keeping assistance was expected, with HC drivers showing the best lane keeping performance and NA drivers showing the worst lane keeping performance. It was furthermore expected that drivers' lane keeping performance deteriorates during periods with concurrent Arrows Task performance due to their visual distraction from the driving task. This deterioration was expected to be most pronounced for NA drivers (driving without lane keeping assistance), and least pronounced for HC drivers. Thus, an interaction between the level of lane keeping assistance and arrows task was predicted.

---

[7] Several critical situations started or ended with other vehicles pulling in or out from the lay-bys on both sides of the road.

The following two measures were analysed: drivers' mean standard deviation of lateral position and the mean duration of the HC system's maximum steering torque application.

4.2.1 Standard deviation of lateral position

Figure 23 shows drivers' mean standard deviation of lateral position over driving periods without and with concurrent Arrows Task performance in Session 1 and Session 2 as a function of the level of lane keeping assistance (group means and standard deviations can be found in Table F1 in Appendix F). One LDW driver was excluded from the analysis because of extreme values that lay more than 2.5 *SD* above the group means.

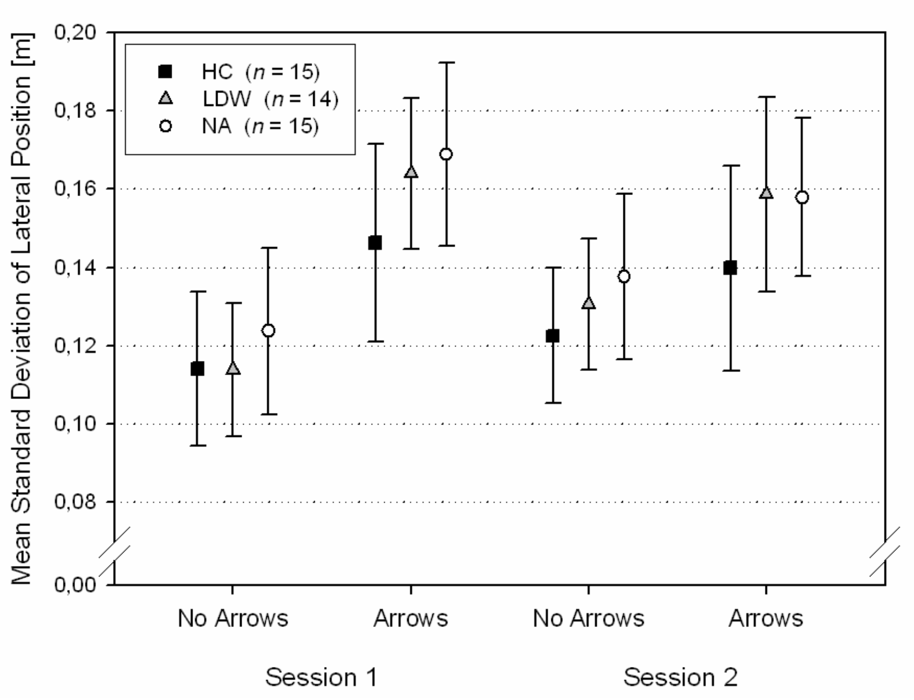


Figure 23. Mean standard deviation of lateral position [m] in periods without and with concurrent Arrows Task performance in Session 1 and in Session 2 as a function of the level of lane keeping assistance. Error bars depict standard deviations.

As predicted, drivers' lane keeping performance clearly deteriorated in driving periods with concurrent Arrows Task performance compared to driving periods without concurrent Arrows Task performance, evidenced by an increase in the standard deviation of lateral position. Figure 23 further provides evidence for the hypothesised improvement in drivers' lane keeping performance as a function of the level of lane keeping assistance. HC drivers tended to show the best lane keeping performance, indicated by the smallest mean standard deviation of lateral position, whereas NA drivers tended to show the worst lane keeping performance, indicated by the largest mean standard deviation of lateral position.

Interestingly, there appeared also to be an unpredicted interaction between the two within-subjects factors session and arrows task. While drivers' lane keeping performance seemed to improve in periods *with* concurrent Arrows Task performance from Session 1 to Session 2, the reverse was found for periods *without* concurrent Arrows Task performance. Actually, drivers' lane keeping performance deteriorated in periods without concurrent Arrows Task performance from Session 1 to Session 2.

A 3 (level of lane keeping assistance) x 2 (session) x 2 (arrows task) repeated-measures ANOVA with session and arrows task as within-subjects factors and level of lane keeping assistance as between-subjects factor was conducted in order to reveal whether any of these effects were statistically significant.

The ANOVA revealed a marginally significant effect of the level of lane keeping assistance, $F(2, 41) = 3.18$, $p = .052$, $\eta^2_G = .10$. There was, as predicted, a highly significant main effect of arrows task, $F(1, 41) = 218.45$, $p < .001$, $\eta^2_G = .37$ (corresponding to a extremely large effect). Furthermore, the interaction between session and arrows task became significant, $F(1, 41) = 40.24$, $p < .001$, $\eta^2_G = .06$ (corresponding to a small effect). Post-hoc comparisons revealed that drivers' lane keeping performance was significantly worse when they concurrently performed the Arrows Task while driving compared to driving alone; both in Session 1, $F(1, 41) = 187.15$, $p < .001$, $\eta^2_G = .52$, and in Session 2, $F(1, 41) = 92.50$, $p < .001$, $\eta^2_G = .22$. Furthermore, it was found that drivers' lane keeping performance significantly deteriorated in periods *without* concurrent Arrows Task performance from Session 1 to Session 2, $F(1, 41) = 22.74$, $p < .001$, $\eta^2_G = .11$. On the contrary, drivers' lane keeping performance significantly improved from Session 1 to Session 2 during periods with concurrent Arrows Task performance, $F(1, 41) = 9.32$, $p = .004$, although this effect was quite small ($\eta^2_G = .03$).

In accordance with the hypotheses, the ANOVA also revealed a significant interaction between the level of lane keeping assistance and arrows task, $F(2, 41) = 3.60$, $p = .04$, $\eta^2_G = .01$. Post-hoc comparisons revealed however that the deterioration in drivers' lane keeping performance from periods without Arrows Task to periods with concurrent performance of the Arrows Task was only significantly larger for LDW drivers than for HC drivers, $F(1, 41) = 7.14$, $p = .01$, $\eta^2_P = .15$[8]. NA drivers did not show a significantly larger deterioration in lane keeping performance due to the concurrent performance of the Arrows Task than HC drivers, $F(1, 41) = 2.28$, $p = .14$, $\eta^2_P = .05$, or than LDW drivers, $F(1, 41) = 1.41$, $p = .24$, $\eta^2_P = .03$.

There was neither a significant main effect of session, $F(1, 41) = 1.80$, $p = .19$, $\eta^2_G = .00$, nor a significant interaction between the level of lane keeping assistance and session, $F(2, 41) = .47$, $p = .63$, $\eta^2_G = .00$. There was also no significant three-way interaction between the level of lane keeping assistance, session and arrows task, $F(2, 41) = .86$, $p = .43$, $\eta^2_G = .00$.

---

[8] η^2_P = partial eta squared.

4.2.2 Mean duration of maximum HC steering torque

As a second measure of drivers' lane keeping performance, the mean duration of the HC system's maximum steering torque application was analysed. Although only drivers in the HC group actually experienced the HC system's steering torques, the forces that would have been generated by the HC system if drivers were assisted by it were also recorded for LDW drivers and for NA drivers. In addition to the analysis of the standard deviation of lateral position as a measure of the magnitude of drivers' fluctuation in lateral position, the mean duration of maximum HC steering torque provides a measure of the time that drivers drove close to the lane boundaries. The mean duration of maximum HC steering torque was calculated as the average duration of the HC system applying its maximum torque in either direction during the approximately 28 s-long single driving periods used for the analysis of drivers' lane keeping performance (see introduction of chapter 4.2).

Figure 24 presents the mean duration of maximum HC steering torque during driving periods without and with concurrent Arrows Task performance in Session 1 and Session 2 as a function of the level of lane keeping assistance (group means and standard deviations can be found in Table F2 in Appendix F).

The data in Figure 24 provides clear evidence for the hypothesised improvement in drivers' lane keeping performance as a function of the level of lane keeping assistance. The mean time during which the HC system would have or had produced its maximum steering torque was largest for NA drivers and shortest for HC drivers. Furthermore, as predicted, Figure 24 shows that drivers' lane keeping performance deteriorated in driving periods with concurrent Arrows Task performance compared to periods of driving alone, indicated by an increase in the mean duration of the HC system's maximum steering torque application. Similarly to the results for the standard deviation of lateral position, the impairment in drivers' lane keeping performance from periods of driving alone to periods with concurrent Arrows Task performance seemed to be more pronounced for Session 1 than for Session 2.

The mean duration of maximum HC steering torque was analysed with a 3 (level of lane keeping assistance) x 2 (session) x 2 (arrows task) repeated-measures ANOVA, with the level of lane keeping assistance as between-subjects factor and session and arrows task as within-subjects factors. Because of the non-normal distribution of the data and heterogeneous error variances in Session 2, the data was additionally analysed with a nonparametric analysis of variance (Kruskal-Wallis *H*) based on aligned ranks (Bortz, Lienert, & Boehnke, 2008). A summary of the results of the parametric and nonparametric ANOVAs is presented in Table 8. The nonparametric ANOVA basically yielded the same results as the parametric ANOVA, except that the main effect of session became significant and that the interaction between session and the level of lane keeping assistance became marginally significant in the nonparametric ANOVA, while both effects were not significant in the parametric ANOVA.

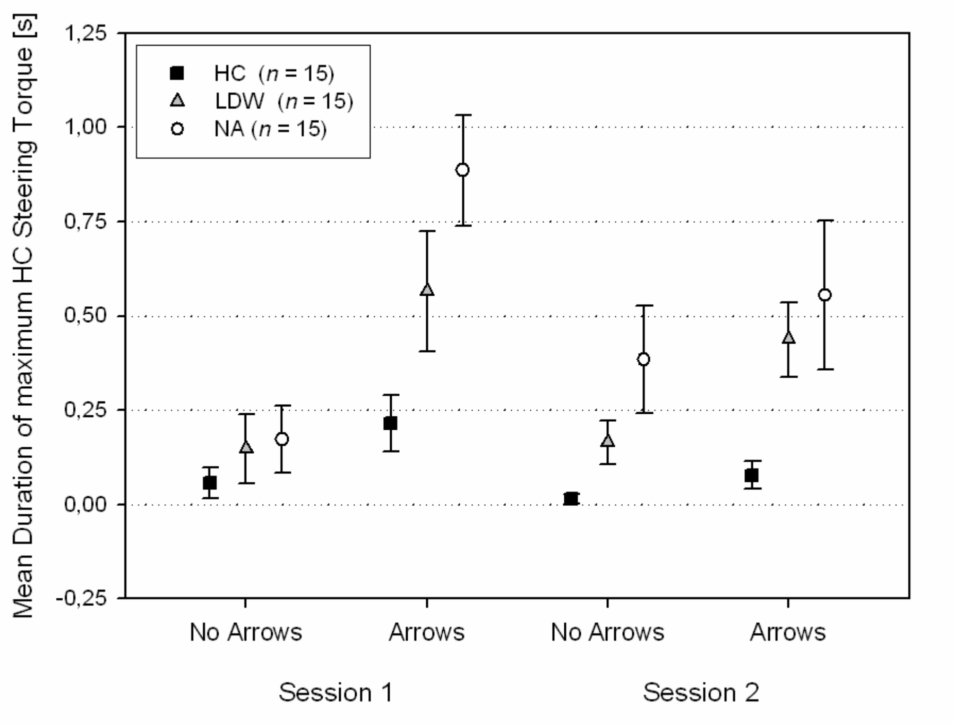


Figure 24. Mean duration of the HC system's maximum steering torque application during the approximately 28 s-long driving periods without and with concurrent Arrows Task performance in Session 1 and Session 2 as a function of the level of lane keeping assistance. Error bars depict standard errors.

For the nonparametric ANOVA, all main effects and interactions except the interaction between session and the level of lane keeping assistance became significant, as can be seen in Table 8. Post-hoc analyses of simple effects involving the between-subjects factor were conducted with nonparametric ANOVAs (Kruskal-Wallis) followed by pairwise comparisons using the Mann-Whitney *U* test. Interaction comparisons involving the repeated factors were analysed with parametric procedures (*F*-tests or *t*-tests) based on the finding that the overall parametric and nonparametric ANOVAs had yielded comparable results.

Simple effects analyses of the between-subjects factor (level of lane keeping assistance) revealed that the three groups of drivers differed significantly in their mean duration of maximum HC steering torque in driving periods with concurrent Arrows Task performance in Session 1, $H(2) = 13.10$, $p = .001$, as well as in driving periods without and with concurrent Arrows Task performance in Session 2, $H(2) = 16.38$, $p < .001$ and $H(2) = 10.41$, $p = .001$, respectively. The effect of the level of lane keeping assistance became only marginally significant in periods without concurrent Arrows Task performance in Session 1, $H(2) = 5.86$, $p = .054$. Pairwise contrasts (Mann-Whitney *U* tests) with Bonferroni-corrected alpha-levels were conducted in order to locate the sources of the significant level of lane keeping assistance effect for the three relevant combinations of the two within-subjects factors (see Table 9). Those pairwise contrasts revealed that in driving periods with concurrent Arrows Task performance in Session 1 HC drivers had a significantly better lane keeping performance than NA drivers, but not than LDW drivers. In Session 2, HC drivers showed a significantly better lane keeping performance than NA drivers and than LDW drivers, both in driving periods without and in driving periods with concurrent Arrows Task performance. Neither in Session 1 nor in Session 2 LDW drivers showed a significantly better lane keeping performance than NA drivers.

Table 8. Results of the parametric and nonparametric analyses of variance for the mean duration of the HC system's maximum steering torque application.

	parametric				nonparametric	
Source	*df*	*F*	η^2_G	*p*	*H*	*p*
			Between subjects			
Level of lane keeping assistance (LoA)	2	5.19*	.14	.01	15.51**	.00
S within-group error	42	(0.49)				
			Within subjects			
Session (Se)	1	2.42	.01	.13	8.48**	.00
Se x LoA	2	.06	.00	.94	5.95	.05
Se x S within-group error	42	(0.09)				
Arrows task (A)	1	34.80**	.13	.00	44.17**	.00
A x LoA	2	3.76*	.03	.03	8.90*	.01
A x S within-group error	42	(0.12)				
Se x A	1	22.47**	.02	.00	55.05**	.00
Se x A x LoA	2	6.58**	.01	.00	43.70**	.00
Se x A x S within-group error	42	(0.03)				

Note. Values enclosed in parentheses represent mean square errors. S = subjects.
$^*p < .05$. $^{**}p < .01$.

Separate ANOVAs for Session 1 and Session 2 confirmed that drivers' lane keeping performance significantly deteriorated in driving periods with concurrent Arrows Task performance compared to periods of driving alone, both in Session 1, $F(1, 42) = 39.42$, $p < .001$, $\eta^2_G = .22$, and in Session 2, $F(1, 42) = 14.38$, $p < .001$, $\eta^2_G = .04$. The interaction between the level of lane keeping assistance and arrows task was significant in Session

1, *F*(2, 42) = 5.47, *p* = .01, η^2_G = .07, but not in Session 2, *F*(2, 42) = 1.89, *p* = .16, η^2_G = .01. Planned contrasts conducted for Session 1 revealed that the increase in the mean duration of maximum HC steering torque from periods of driving alone to driving periods with concurrent Arrows Task performance was significantly larger for LDW drivers than for HC drivers, *t*(22.16) = 2.18 (adjusted for unequal variances among groups), *p* = .04 (two-tailed), *r* = .38. NA drivers did not show a significantly larger increase in the mean duration of HC maximum steering torque than LDW drivers in Session 1, *t*(23.26) = 1.50 (adjusted for unequal variances among groups), *p* = .15 (two-tailed), *r* = .27.

Table 9. Results of pairwise contrasts (Mann-Whitney *U* tests) between the three levels of lane keeping assistance for each combination of the two within-subjects factors session and arrows task.

	Session 1				Session 2			
	No Arrows		Arrows		No Arrows		Arrows	
Contrast	*Z*	*p*[a]	*Z*	*p*[a]	*Z*	*p*[a]	*Z*	*p*[a]
HC - NA	[b]		-3.38*	.000	-3.42*	.000	-2.77*	.006
HC - LDW	[b]		-2.10	.035	-3.79*	.000	-2.83*	.004
LDW - NA	[b]		-1.93	.054	-1.06	.300	-0.35	.743

[a] two-tailed. [b] no pairwise contrasts were conducted because of a non-significant result of the omnibus ANOVA (Kruskal-Wallis).
*significant after Bonferroni-correction for multiple comparisons.

There was no significant increase in the mean duration of maximum HC steering torque in driving periods without concurrent Arrows Task performance from Session 1 to Session 2, *F*(1, 42) = 3.44, *p* = .07, η^2_G = .01. However, there was a significant interaction between the level of lane keeping assistance and session, *F*(2, 42) = 5.19, *p* = .01, η^2_G = .03. Thus, while HC drivers showed a slightly better lane keeping performance in driving periods without concurrent Arrows Task performance in Session 2 compared to Session 1, LDW drivers' and NA drivers' lane keeping performance deteriorated in driving periods without concurrent Arrows Task performance from Session 1 to Session 2. Post-hoc comparisons revealed that NA drivers showed a significantly larger increase in the mean duration of maximum HC steering torque than HC drivers from Session 1 to Session 2 in periods of driving alone, *F*(1, 28) = 7.69, *p* = .01, η^2_G = .01.

There was further a significant main effect of session when drivers' lane keeping performance was compared in periods with concurrent Arrows Task performance in Session 1 and Session 2. Thus, drivers' lane keeping performance significantly improved in periods with concurrent Arrows Task performance from Session 1 to

Session 2, $F(1, 42) = 9.33$, $p = .004$, $\eta^2_G = .04$. The magnitude of improvement was not significantly different for the three levels of lane keeping assistance, $F(2, 42) = .36$, $p = .004$, $\eta^2_G = .01$.

4.2.3 Summary and discussion

Drivers' perception that the lane keeping assistance systems improved their lane keeping performance was regarded as a prerequisite for the hypothesised changes in drivers' active engagement in the driving task as a function of the level of automation of the lateral control task. In order to check whether drivers' lane keeping performance actually improved as a function of the increasing automation of the lateral control task, two measures of drivers' lane keeping performance were analysed: the mean standard deviation of lateral position as a measure of the magnitude of drivers' variation in lateral position, and the mean duration of the HC system's maximum steering torque application as a measure of the time drivers drove close to the lane boundaries.

Both measures yielded clear evidence that drivers' lane keeping performance improved with the increasing automation of the lateral control task. Drivers' variation in lateral position decreased the higher the level of lane keeping assistance, although this effect became only marginally significant. However, there was a significant effect of the level of lane keeping assistance on the time that drivers drove close to the lane boundaries, as indicated by the mean time during which the HC system had or would have applied its maximum steering torque. Except for driving periods without concurrent Arrows Task performance in Session 1, the mean duration of the HC system's maximum steering torque application was significantly shorter for HC drivers than for drivers driving without lane keeping assistance (NA drivers), both in driving periods without and in driving periods with concurrent Arrows Task performance. Furthermore, the mean duration of the HC system's maximum steering torque application was also significantly shorter for HC drivers than for LDW drivers in Session 2. Although NA drivers drove a higher amount of time close to the lane boundaries than LDW drivers, the difference between both groups was not significant. However, the fact that LDW drivers showed a better lane keeping performance than NA drivers indicates that the LDW system's warnings encouraged LDW drivers to invest more effort in maintaining a good lane keeping performance, probably in order to minimise the number of lane departure warnings.

As predicted, drivers' lane keeping performance significantly deteriorated when drivers had to perform a visually demanding secondary task concurrently to driving. It was hypothesised that this effect would be most pronounced for NA drivers and least pronounced for HC drivers as a function of the increasing automation of the lateral control task. Again, evidence for this prediction was only found for the mean duration of the HC system's maximum steering torque application, and not for the mean standard deviation of lateral position. Furthermore, the effect was only evident in Session 1, with NA drivers and LDW drivers showing a significantly larger increase in the time they drove close to the lane boundaries than HC drivers from driving periods without concurrent Arrows Task performance to driving periods with concurrent

Arrows Task performance. The finding that this effect was absent in Session 2 can be interpreted as a sign for drivers' increased attempt to protect their primary task performance from safety-critical deteriorations due to their concurrent performance of the Arrows Task in Session 2.

Further evidence for this interpretation of results is provided by the finding that drivers' lane keeping performance significantly improved in periods *with* concurrent Arrows Task performance from Session 1 to Session 2. This effect was significant for both measures of lane keeping performance. Thus, the unexpected occurrence of critical driving situations in periods with concurrent Arrows Task performance in Session 2 apparently provoked drivers to invest more effort in lane keeping in an attempt to increase their safety margins by driving closer to the centre of the lane.

Interestingly, the opposite result was found for driving periods *without* concurrent Arrows Task performance in Session 1 and Session 2. There was a significant increase in the standard deviation of lateral position from Session 1 to Session 2 in periods of driving alone. This increase in lateral position variation did however not result in a general increase in the time drivers drove close to the lane boundaries. Only NA drivers were found to show a significantly larger increase in the mean duration of the HC system's maximum steering torque application than HC drivers in periods without concurrent Arrows Task performance from Session 1 to Session 2.

Altogether, these results speak for a shift in drivers' mental effort regulation strategies from Session 1 to Session 2. It seems that drivers took more efforts to protect their primary task performance in driving periods *with* concurrent Arrows Task performance in Session 2 that were especially vulnerable to safety-critical deteriorations in lane keeping performance because of the unexpected occurrence of critical driving situations. Part of these critical situations required drivers to react on the lateral control level by drawing aside to either the left or to the right lane boundary in order to avoid potential collisions with other traffic. Thus, driving close to the centre of the lane constituted probably the best strategy for drivers to be able to react in an appropriate and timely manner to these situations. The significant improvement in drivers' lane keeping performance in periods with concurrent Arrows Task performance from Session 1 to Session 2 can thus be interpreted as an active attempt of drivers to increase their safety margins in light of the increasing safety and task demands in Session 2. On the other hand, the significant impairment in drivers' lane keeping performance in periods *without* concurrent Arrows Task performance from Session 1 to Session 2 appears to indicate that drivers tried to compensate for the increased effort investment in periods with concurrent Arrows Task performance in Session 2. Thus, the deterioration in drivers' lane keeping performance under single-task (driving alone) conditions from Session 1 to Session 2 may be attributed to an attempt of drivers to cope with the increasing driving task demands in Session 2 by lowering primary task performance standards in driving periods that were less vulnerable to deteriorations in lane keeping performance. Mental effort regulation theories (Hockey, 1997) predict that operators switch to less resource-intensive performance strategies after periods of increased effort investment when task demands permit it.

4.3 Drivers' Reliance on the Lane keeping Assistance Systems

Drivers' strategies for allocating their visual attention between the road scene and the visual display of the secondary task (Arrows Task) served as a measure for their reliance on the lane keeping assistance systems. It was assumed that drivers' increasing reliance on the lane keeping assistance systems would be reflected by their increased preparedness to allocate their visual attention away from the road scene to the visual display of the Arrows Task. Two measures of drivers' visual attention allocation strategies were assumed to be sensitive to different levels of reliance, namely the *mean duration of drivers' single glances to the Arrows Task display*, and the *mean proportion of display glance time* (i.e., the proportion of time that drivers glanced to the in-vehicle visual display until they responded to the single displays of the Arrows Task). In the following section it is explained how these two measures were derived from the raw eye movement data recorded by the eye tracking system.

4.3.1 Analysis of drivers' eye glance behaviour

The two measures of reliance – the mean duration of drivers' single glances to the Arrows Task display and the mean proportion of display glance time – were extracted by analysing the raw data of drivers' gaze direction recorded by the eye tracking system. Drivers' gaze direction was measured as the x- (left-right), y- (up-down) and z- (near-far) coordinates of a unit vector at a frame rate of 60 Hz. Thus, no areas of interest were defined a priori. A first glance at the eye movement data revealed however that there was a substantial variation in tracking quality between drivers, and even within individual drivers (e.g., between Session 1 and Session 2). In order to minimise the loss of data due to imprecision in tracking, manually controlled procedures were used for the analysis of drivers' gaze direction instead of automated routines.

The first step in the analysis consisted of plotting the relevant eye movement parameters (x-, y- and z- coordinates of drivers' gaze direction, gaze tracking quality, and head rotation) for each of the 44 Arrows Tasks that every driver performed during the course of the study (Plot A, see Figure 25)[9]. An in-depth inspection of these plots revealed that the eye movement data of four drivers (three drivers in the LDW group

---

[9] Drivers performed 19 Arrows Tasks during the first experimental session and 25 Arrows Tasks during the second experimental session. Each Arrows Task had a duration of 30 s.

and one driver in the control group) were not suitable for further processing because of an insufficient tracking quality.

The eye tracking data of the remaining 41 drivers was analysed in the following steps:

1. A second plot (Plot B) was created for all remaining Arrows Tasks (*N* = 1791)[10] which displayed the sum of the x-, y-, and z- coordinates of the left eye's (in blue) and of the right eye's (in red) gaze vector. This additional plot was created to yield a greater separation between drivers' glances to the road scene and glances to the Arrows Task display (see Figure 26). Glances to the road were defined as glances to areas any other than to the Arrows Task display.
2. For each Plot B, a threshold was defined manually that was set to discriminate best between drivers' glances to the road and their glances to the Arrows Task display. The threshold was set either for the right eye or for the left eye depending on which eye's parameters were better suitable for discrimination of the two gaze directions (see Figure 26). The threshold was set by one observer, and afterwards verified by two independent observers in order to ensure that no misjudgements occurred. In the case of difficulties regarding the classification of single glances (e.g., in terms of their starting point, especially when tracking quality was low), the respective scenes in the video recordings of the drivers' faces were observed in order to verify the observer judgements. When particular glances could not be properly classified according to the manually defined threshold, those instances (periods of time) were recorded and later (manually) excluded from analysis. Exclusion of data was based on complete glances in order to ensure that the measure of single glance duration was not affected by this procedure.
3. A custom-written Matlab-algorithm was applied to the eye movement data of each driver that extracted drivers' single glances to the road and to the Arrows Task display based on the threshold defined manually in step 2. The algorithm extracted the starting times and end times of drivers' glances, as well as the duration of these glances.
4. The method of setting the threshold (step 2) based on the sum of the x-, y-, and z-coordinates of the gaze vector (presented in Plot B) could actually lead to an incorrect classification of drivers' glances to the rear mirror, the outside mirrors and also to the tachometer. Those glances could be incorrectly classified as glances to the Arrows Task display because the summation of the gaze vector coordinates yielded values that were comparable to those originating from actual glances to the Arrows Task display (see Figure 26). For this reason, the glances extracted by the algorithm (in step 3) were verified again by the three independent observers by comparing them with the plots of the "raw" coordinates of the gaze vectors (Plot A) where glances to the rear mirror and the outside mirrors could easily be identified through a dissociation of the x-

---

[10] 41 drivers multiplied by 44 Arrows Task minus 13 Arrows Tasks (due to technical problems, the eye movements of one driver were not recorded during 13 Arrows Tasks in Session 2).

and y- gaze coordinates. In the case of an incorrect classification of glances, a correction algorithm was applied that appropriately reclassified glances to the rear or outside mirror as non-display glances (i.e., as glances to the road).

5. A second correction algorithm was applied that searched for potentially incorrectly classified glances (e.g., glances with a duration shorter than 200 ms, instances of two subsequent glances to the same direction).
6. In a final step, drivers' extracted glances to either the road or the Arrows Task display were synchronised with the data from the Arrows Task (recorded in a separate file). By doing this, the presentation times of the single displays of one Arrows Task were matched with the starting and end times of drivers' glances to the road and to the Arrows Task display. The synchronisation was necessary to be able to exclude display glances from the analysis that occurred *between* the presentations of two subsequent Arrows Task displays. Those "check" glances were attributable to drivers' inspection of the performance feedback that appeared on the screen after they gave their response to the Arrows Task. Furthermore, the synchronisation was required for the calculation of the proportion of display glance time, which was defined as the proportion of time that drivers glanced to the Arrows Task display during the presentation of a new Arrows Task display and drivers' responses to that display (see Figure 27).

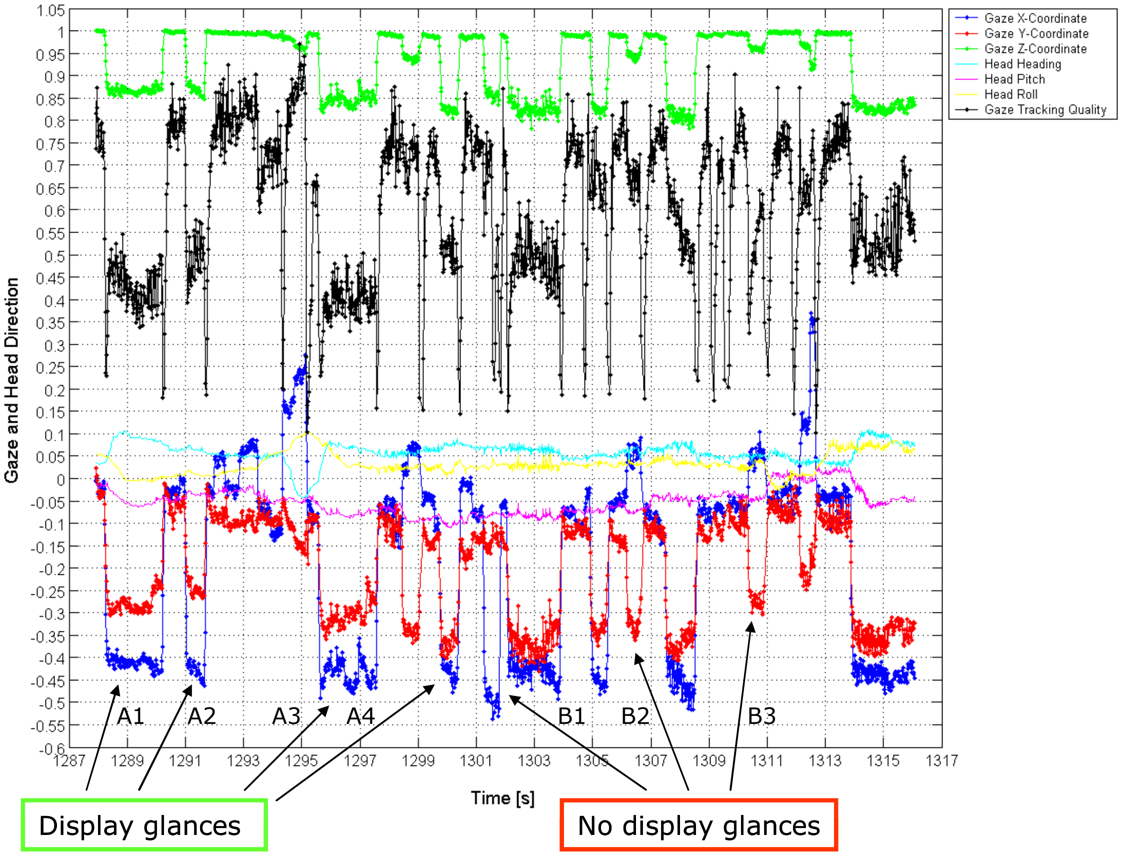


Figure 25. Plot A: Example of one driver's gaze and head direction during one Arrows Task. A1 to A4 represent glances to the Arrows Task display. B1 to B3 represent glances to the right outside mirror and to the tachometer (identifiable through a dissociation of the x- and y- direction of gaze).

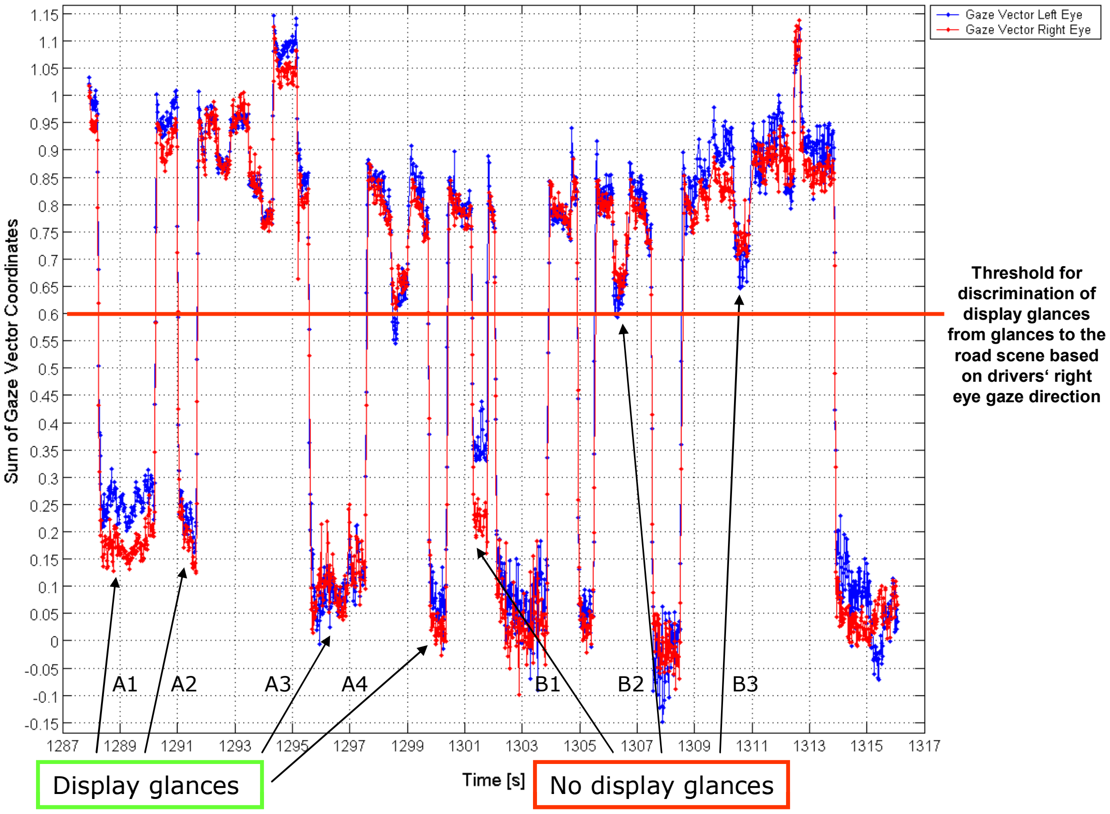


Figure 26. Plot B: Sum of gaze vector coordinates (x, y, and z) measured for the driver's left eye (in blue) and the driver's right eye (in red) during the Arrows Task displayed in Plot A. The red line represents the manual threshold defined for the discrimination of glances to the Arrows Task display (below the threshold) from glances to the road scene (above the threshold) based on the data recorded for the driver's right eye. B2 and B3 are correctly classified as "no display glances", whereas B1 is incorrectly classified as a display glance by the algorithm and therefore corrected in a later step of the analysis.

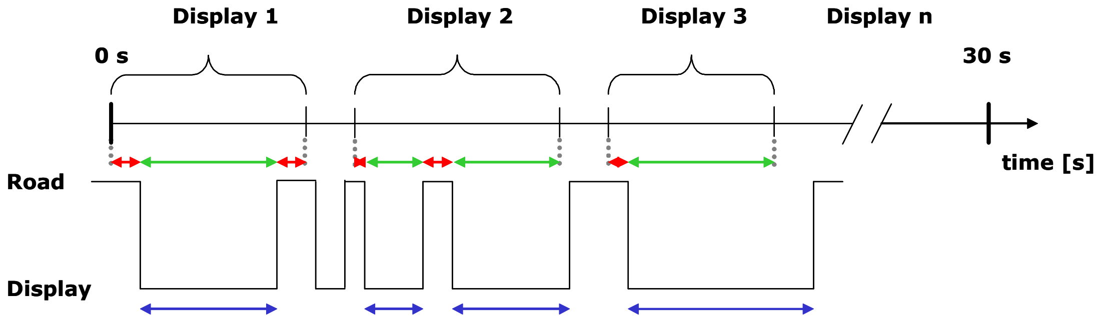


Figure 27. Calculation of the mean proportion of display glance time and of the mean single display glance duration for one Arrows Task (30 s). The presentation times of the single Arrows Task displays (1 to n) equal drivers' response times to the single displays. The mean proportion of display glance time is the average proportion of time that a driver glanced to the Arrows Task display during the presentation of the single displays (represented by the green arrows). The mean single display glance duration is the average duration of drivers' single glances to the Arrows Task display (represented by the blue arrows) excluding glances that occurred between the presentation of two subsequent displays (between display 1 and display 2 in the figure).

4.3.2 Subsequent analysis of reliance measures

Drivers' reliance on the lane keeping assistance systems was investigated as a function of the level of lane keeping assistance (between-subjects factor) and as a function of drivers' experience of functional system limits (within-subjects factor). Drivers' experience of functional system limits was manipulated by help of the session factor: Whereas Session 1 consisted only of "normal" driving periods (without occurrence of critical driving situations), Session 2 included normal driving periods as well as eight critical driving situations (of which three simulated functional limits of the lane keeping assistance systems by requiring drivers to react on the lateral control level).

It was hypothesised that drivers' experience of functional system limits in Session 2 would lead to a decrease in their reliance on the lane keeping assistance systems. Thus, in response to the experience of functional system limits, drivers were assumed to be less prepared to allocate their visual attention away from the road scene to the visual display of the Arrows Task. This decrease in drivers' reliance from Session 1 to Session 2 was assumed to become apparent *not only at the very instance of drivers' encountering of critical situations, but also during normal driving periods* as a result of drivers' comprehension that they could not rely under any circumstances on the lane keeping assistance systems' support. Thus, a decrease in drivers' reliance should also become apparent in normal driving situations in Session 2 as a consequence of drivers' *anticipation* that they may encounter occasions in which they could not rely on the support provided by the lane keeping assistance systems (note that drivers did neither know *that* they would encounter critical driving situations in Session 2, nor *when* and *how many* of them would occur).

In order to test the hypotheses about the effect of drivers' experience of functional system limits on their reliance on the lane keeping assistance systems, measures of drivers' reliance were analysed over three levels of the repeated measures factor: (a) over normal driving periods in Session 1, (b) over normal driving periods in Session 2, and (c) in critical driving situations (Events) in Session 2. Thus, the design was a 3 (level of lane keeping assistance: high, low, none) x 3 (session-event: Session 1, Session 2, Events) mixed factorial design.

The durations of single display glances and the proportions of display glance time were first aggregated over six consecutive Arrows Task sections for each driver (each consisting of six 30 s-Arrows Tasks *excluding* those where critical driving situations occurred). The first three Arrows Task sections belonged to Session 1, whereas the second three Arrows Task sections (four to six) belonged to Session 2. The sections four to six of Session 2 included Arrows Tasks that occurred *after* drivers had encountered one up to two critical driving situations. In order to derive the third level of the within-subjects factor, the durations of single display glances and the proportions of display glance time were aggregated over all Arrows Tasks that occurred during critical driving situations. After that, extreme values for the two reliance measures that lay more than 2.5 *SD* above or below the respective means of the seven Arrows Task sections (three in Session 1, three in Session 2, and one comprising all Arrows Tasks where critical situations occurred) for each driver were removed from the data. After removal of the outliers, the mean duration of single display glances and the mean proportion of display glance time of each driver for Session 1 were derived from aggregating Arrows Task sections one to three, and the respective means for normal driving situations in Session 2 were derived from aggregating Arrows Task sections four to six.

4.3.3 Mean duration of drivers' single glances to the Arrows Task display

For the analysis of the mean single display glance duration, glances that occurred between drivers' responses to one Arrows Task display and the presentation of the next display were excluded from the analysis ("check" glances that could presumably be attributed to drivers' inspection of the visual performance feedback that was displayed following their responses to the Arrows Task). Those check glances were short in duration and were considered as not belonging to the task.

Due to insufficient gaze tracking quality, the eye glance behaviour of four drivers (three drivers in the LDW group and one driver in the control group) could not be analysed. After the exclusion of the data of these four drivers, a total number of 12,348 display glances were included in the analysis of the mean duration of drivers' single glances to the Arrows Task display. Due to the removal of outliers from the data of each individual driver (procedure explained above), 1.1% of the glances were excluded from the analysis.

Figure 28 shows the mean duration of drivers' single glances to the Arrows Task display as a function of the level of assistance over the three levels of the repeated measures factor (session-event). The mean durations depicted for each of the three levels of the session-event factor (x-axis) are derived from about 128 display glances for each driver in Session 1, from about 115 display glances for each driver in normal driving periods in Session 2, and from about 50 glances for each driver in critical situations (Events) in Session 2.

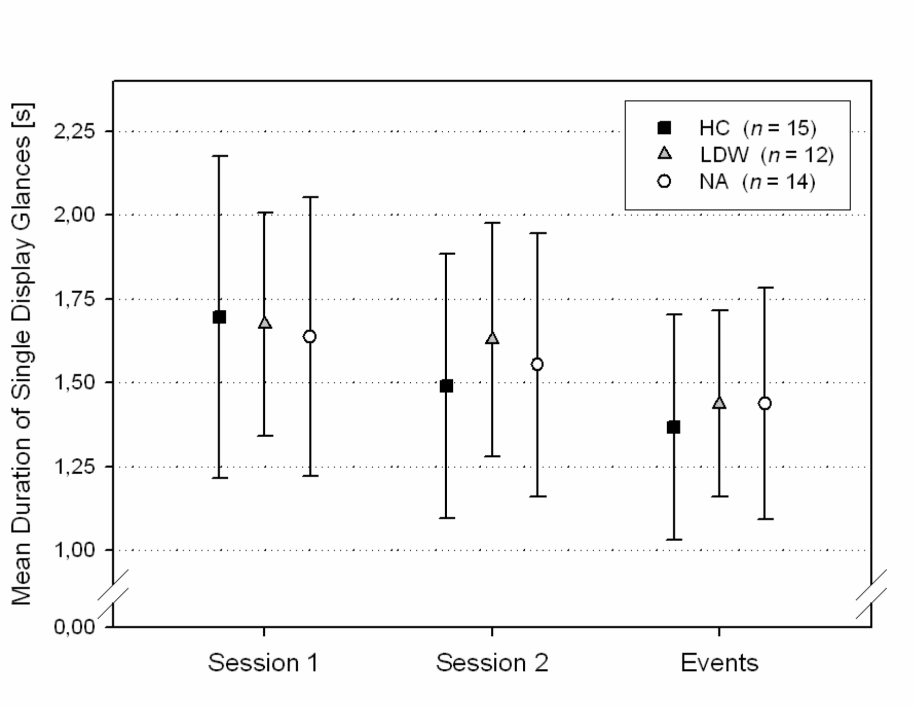


Figure 28. Mean duration of drivers' single glances [s] to the Arrows Task display as a function of the level of lane keeping assistance in Session 1 (consisting of normal driving periods only), in normal driving periods in Session 2 and in critical driving situations in Session 2 (Events). The linear decrease from Session 1 over Session 2 to critical driving situations is highly significant (r = .28). Error bars depict standard deviations.

It was predicted that drivers' experience of critical driving situations in Session 2 (partly representing functional limits of the LDW and HC system) would have a negative effect on drivers' reliance on the lane keeping assistance systems. As a response to the unexpected occurrence of critical driving situations in Session 2 drivers were assumed to be less prepared to allocate their visual attention away from the road scene for longer periods of time, reflected by a decrease in the mean duration of their single glances to the Arrows Task display. This decline in drivers' mean single display glance durations from Session 1 (where no critical driving situations occurred) to Session 2 should not only result from shorter eyes-off-the-road times *when drivers actually encountered critical situations*. Rather, it was predicted that drivers reduce their eyes-off-the-road times *also during normal driving periods in Session 2* as a response to their *understanding* that they could not rely under any circumstances on the lane keeping assistance systems because they were providing inadequate support in some situations. Thus, drivers' mean single display glance durations should decrease from Session 1 to normal driving periods in Session 2, and further decrease from normal driving periods in Session 2 to critical driving situations (Events) in Session 2.

As Figure 28 shows, this predicted linear decrease over the three levels of the session-event factor was clearly supported by the data, and proved to be highly significant, $t(38) = 7.32$, $p < .001$ (one-tailed), $r = .28$. Pairwise contrasts revealed that the effect was slightly larger for the decline in single display glance duration from normal driving periods in Session 2 to critical driving situations in Session 2, $t(38) = 6.81$, $p < .001$ (one-tailed), $r = .19$, than for the decline from normal driving periods in Session 1 to normal driving periods in Session 2, $t(38) = 3.95$, $p < .001$ (one-tailed), $r = .15$.

A second major hypothesis concerned the effect of the level of lane keeping assistance on drivers' reliance in Session 1 *before* they experienced functional limits of the lane keeping assistance systems. It was hypothesised that drivers are prepared to allocate increasing levels of control to a lane keeping assistance system (reflected by their higher preparedness to allocate their visual attention away from the road scene) the more completely it automates the lateral control task. Thus, the reduced need to actively intervene in lateral control with higher levels of lane keeping assistance should result in drivers' less active engagement in the driving task and their allocation of a higher amount of control to the lane keeping assistance systems for maintaining an acceptable driving performance.

There is no clear indication in Figure 28 that HC drivers had longer mean single display glance durations than LDW and NA drivers in Session 1. However, the histograms of the distributions of drivers' single display glance durations for the three levels of lane keeping assistance revealed an interesting finding (see Figure 29 and Figure 30). Unlike the glance distributions for LDW and NA drivers in Session 1, the glance distribution for HC drivers in Session 1 seemed to be composed actually of two overlapping distributions, one with a mean around 1.3 s, and one with a mean around 2 s. Exploratory data analysis confirmed this finding that the HC group of drivers could be divided into two subgroups according to their mean single glance duration over the levels of the session-event factor. Thus, there was one subgroup of HC drivers (consisting of five drivers) that had particularly long mean single display glance durations, and another subgroup of HC drivers (consisting of 10 drivers) that had particularly short mean single display glance durations.

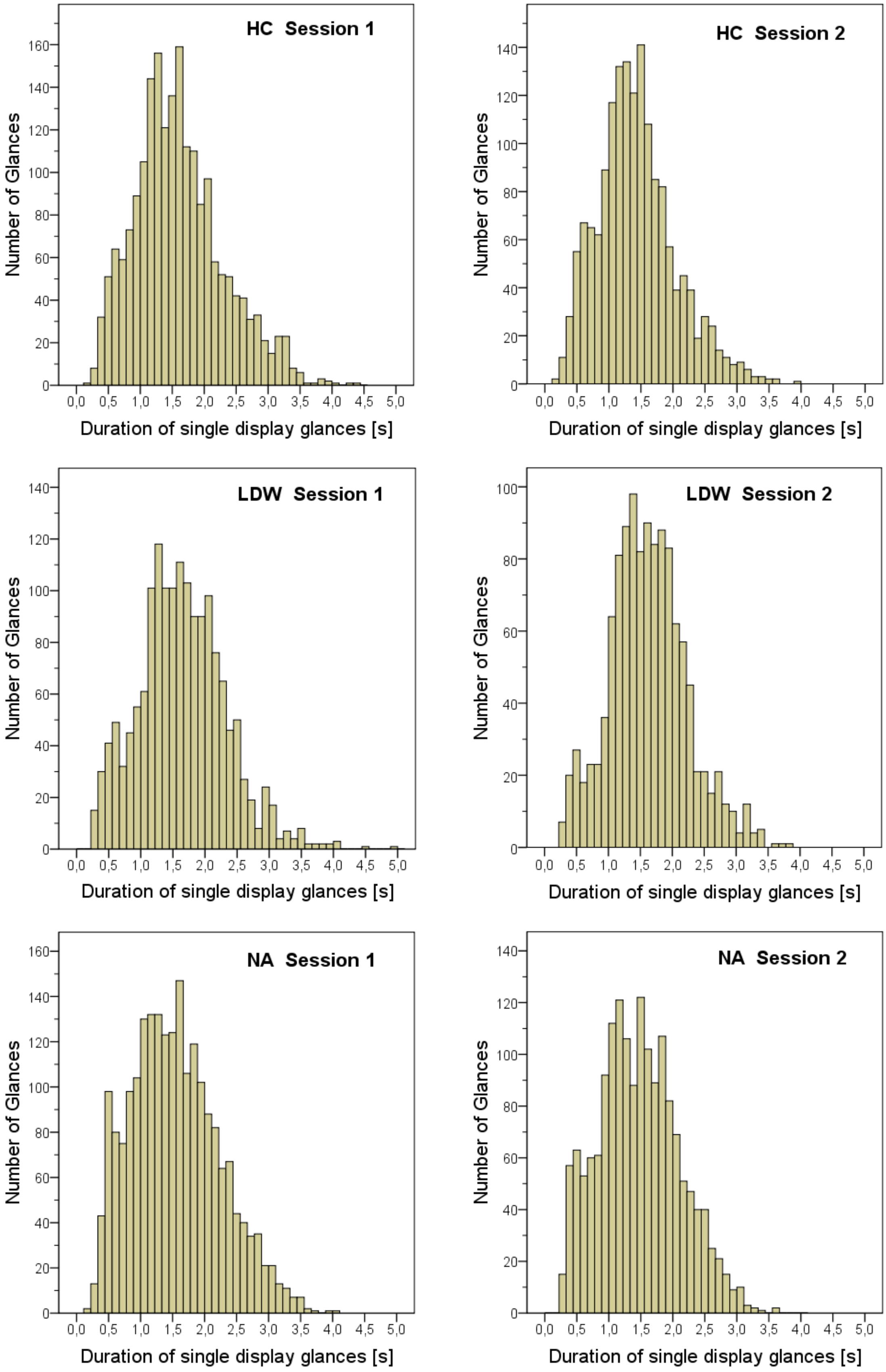


Figure 29. Distribution of drivers' single display glance durations for the three levels of lane keeping assistance in normal driving periods in Session 1 and Session 2.

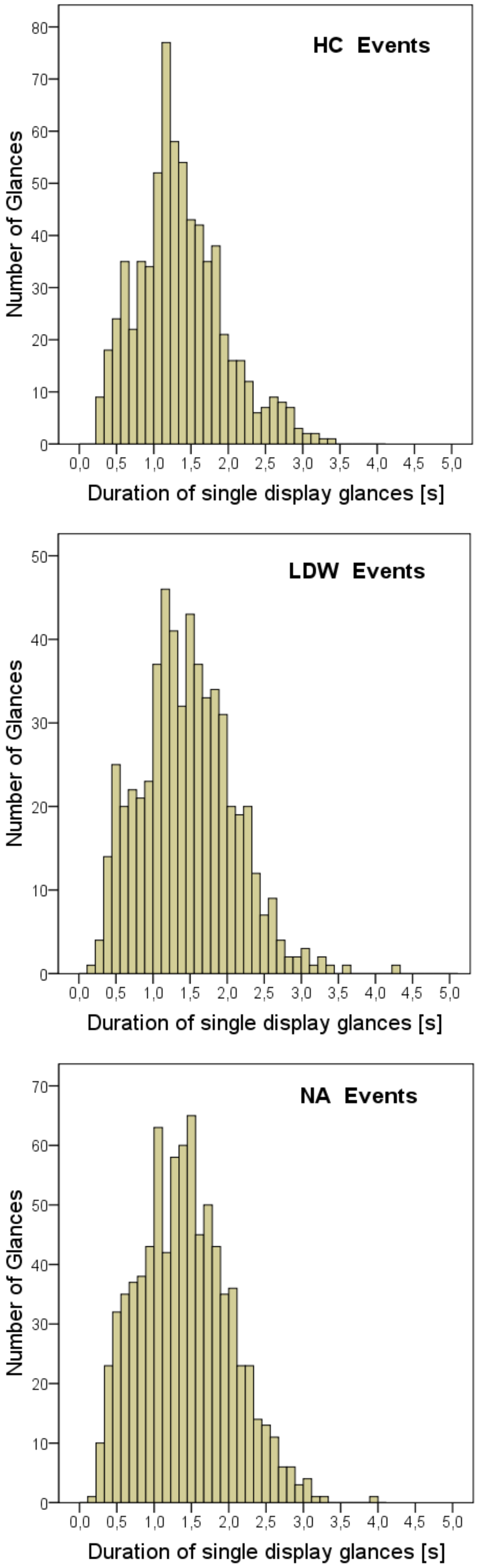


Figure 30. Distribution of drivers' single display glance durations for the three levels of lane keeping assistance in critical driving situations (Events) in Session 2.

Because this appeared to be a clear effect of the manipulation of the level of lane keeping assistance factor, the two subgroups of HC drivers were treated separately in the following analyses. Figure 31 displays the same data as Figure 28, but separately for the two subgroups of HC drivers (referred to as "HC long" and "HC short" in Figure 31). The mean durations of single display glances and standard deviations for the two subgroups of HC drivers and for LDW and NA drivers across the three levels of the session-event factor are presented in Table 10.

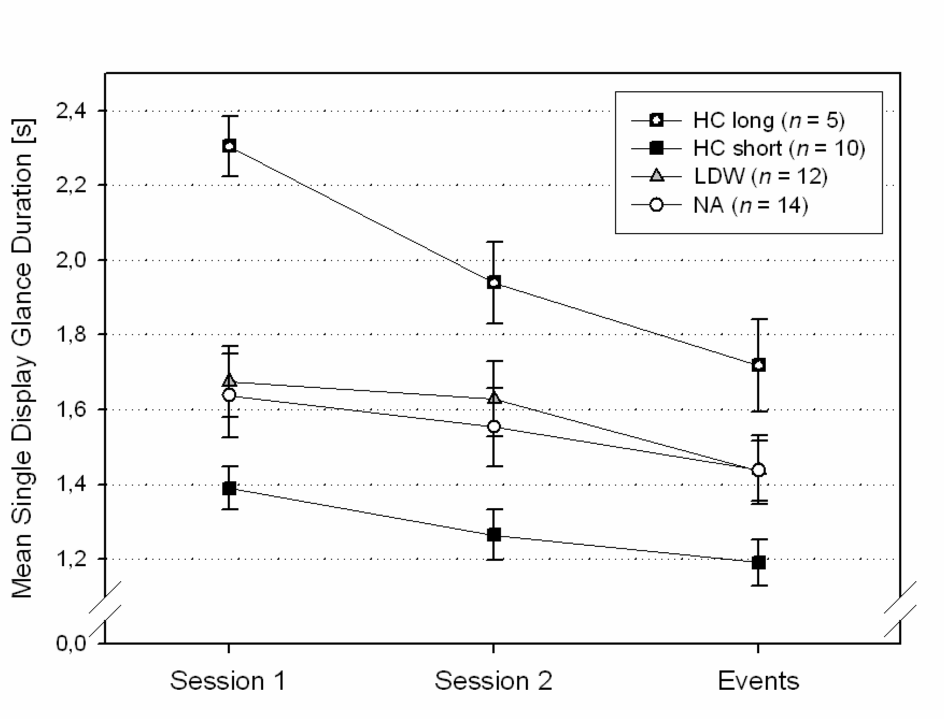


Figure 31. Mean duration of single display glances [s] as a measure of drivers' reliance on the lane keeping assistance systems as a function of the level of assistance. Drivers differed substantially in their reliance on the HC system and could be divided into two subgroups, one subgroup ('HC long') that relied to a high extent on the HC system (indicated by their long mean single glance durations) and a second subgroup ('HC short') that apparently did not rely on the HC system at all. Error bars depict standard errors.

As can be seen in Figure 31, the subgroup of drivers with the long mean single display glance durations ('HC long' subgroup) had substantially longer display glance durations than LDW and NA drivers in Session 1 (in accordance with the hypotheses); however, their glance durations were also longer than those of the LDW and NA drivers in Session 2 and in critical driving situations (Events). This latter finding contradicted the original hypotheses. In fact, it was assumed that HC drivers would have the shortest single display glance durations in Session 2 and in critical driving situations because they should be most strongly affected by the experience of functional system limits due to their need to actively override the HC system.

On the contrary, the subgroup of drivers with the short mean single display glance durations ('HC short' subgroup) had - in accordance with the hypotheses - shorter display glance durations than LDW and NA drivers in Session 2 and in critical driving situations. However, they had also shorter mean display glance durations in Session 1, contrary to the hypothesis.

Apparently, drivers in the HC group reacted very differently to the HC system's support on a general level by allocating either a very high level of control to the system or by not relying on the system at all, resulting in even shorter single glances away from the road scene compared to the control (NA) group.

Table 10. Mean duration of drivers' single glances to the Arrows Task display as a measure of drivers' reliance on the lane keeping assistance systems for the two subgroups of HC drivers and for LDW and NA drivers. Mean durations of drivers' single display glances decrease from Session 1 to normal driving periods in Session 2 as a response to drivers' unexpected encountering of critical driving situations in Session 2, and are shortest during critical driving situations (Events).

Session 1		Session 2		Events	
M	*SD*	*M*	*SD*	*M*	*SD*
'HC long' subgroup (*n* = 5)					
2.30	0.18	1.94	0.25	1.72	0.28
'HC short' subgroup (*n* = 10)					
1.39	0.18	1.26	0.22	1.19	0.20
LDW drivers (*n* = 12)					
1.68	0.33	1.63	0.35	1.44	0.28
NA drivers (*n* = 14)					
1.64	0.42	1.55	0.39	1.44	0.34

In order to test whether the two HC subgroups differed significantly in their mean single display glance durations from LDW and NA drivers, an ANOVA was conducted with the level of lane keeping assistance as between-subjects factor (with four levels: HC long, HC short, LDW, and NA) and session-event as within-subjects factor (with three levels: Session 1, Session 2, Events). Levene-tests confirmed that the variance homogeneity assumption was not violated over the three levels of the session-event factor. However, the sphericity assumption was not met. Therefore ε-adjusted degrees of freedom (lower bound) were used. The ANOVA revealed a significant and large main effect of the level of lane keeping assistance, $F(3, 37) = 6.28$, $p = .001$, $\eta^2_G = .32$. There was also a significant main effect of the within-subjects factor session-event, *F*(1, 37) = 60.14, $p < .001$, $\eta^2_G = .13$, confirming the linear decrease in mean single glance durations from Session 1 over normal driving periods in Session 2 to critical driving situations in Session 2. The Level of lane keeping assistance x Session-Event interaction proved also significant, however, the effect was small in magnitude, *F*(3, 37) = 4.58, $p = .008$, $\eta^2_G = .03$.

Post-hoc comparisons (with Bonferroni-corrected α-levels) revealed a strong and significant effect for the two HC subgroups' difference in mean single display glance duration over all levels of the session-event factor, yielding evidence that the two subgroups of HC drivers differed substantially in their reliance on the HC system (see contrast 3 in Table 11). Post-hoc tests further revealed that both HC subgroups differed significantly from both LDW and NA drivers in Session 1, with drivers in the 'HC long' subgroup glancing significantly longer away from the road scene than LDW and NA drivers, and drivers in the 'HC short' subgroup glancing significantly shorter away from the road scene than LDW and NA drivers (see contrast 1 and 3 in Table 11). In Session 2 and in critical situations, the HC subgroups did not differ significantly in their mean single glance durations to the Arrows Task display from both LDW and NA drivers.

Table 11. Contrasts conducted between levels of the between-subjects factor (level of lane keeping assistance) for each level of the within-subjects factor (session-event) in order reveal the sources of the significant Level of lane keeping assistance x Session-Event interaction.

Session 1		Session 2		Events	
t(37)	*r*	*t*(37)	*r*	*t*(37)	*r*
Contrast 1 ($\lambda_{\text{HC long}} = 2$, $\lambda_{\text{LDW}} = -1$, $\lambda_{\text{NA}} = -1$, $\lambda_{\text{HC short}} = 0$)					
4.09*	.56	2.17	.34	2.01	.31
Contrast 2 ($\lambda_{\text{HC long}} = 0$, $\lambda_{\text{LDW}} = 1$, $\lambda_{\text{NA}} = 1$, $\lambda_{\text{HC short}} = -2$)					
2.21	.34	2.66	.40	2.30	.35
Contrast 3 ($\lambda_{\text{HC long}} = 1$, $\lambda_{\text{LDW}} = 0$, $\lambda_{\text{NA}} = 0$, $\lambda_{\text{HC short}} = -1$)					
5.15*	.65	3.75*	.52	3.35*	.48

Note: * significant after Bonferroni-correction for multiple comparisons.

A third hypothesis concerned the magnitude of the change in drivers' reliance in response to the occurrence of unexpected critical driving situations (partly representing functional system limits). It was predicted that the decline in drivers' reliance from Session 1 to Session 2 would be most pronounced for HC drivers, less pronounced for LDW drivers and least pronounced for NA drivers, as a function of degree of the lane keeping assistance systems' intervention in driving. In order to test this hypothesis, planned contrasts were conducted. The results indicate that only the 'HC long' subgroup showed a significantly larger decrease in reliance than the LDW group as response to their experience of functional system limits, $t(37) = 3.55$, $p < .001$ (one-tailed), $r = .50$, whereas the 'HC short' subgroup did not show a significantly more pronounced decrease in reliance than the LDW group, $t(37) = 1.08$, $p = .14$ (one-tailed), $r = .17$. LDW drivers did not show a larger decline in mean single display glance duration from Session 1 to Session 2 than NA drivers, $t(37) = -.55$, $p = .29$ (one-tailed), $r = -.09$.

4.3.4 Mean proportion of display glance time

The proportion of time drivers glanced to the Arrows Task display since the presentation of an Arrows Task display until they gave their response was used as a second indicator for drivers' reliance on the lane keeping assistance systems. The mean proportion of display glance time was calculated based on Arrows Tasks only for which all glances to the road or the Arrows Task display were available. Arrows Tasks for which single glances could not be accordingly classified (according the procedure explained in section 4.3.1) were excluded from the analysis. Furthermore, the proportion of display glance time was calculated based on Arrows Task displays only that drivers had responded to (excluding displays with missing responses) and for which response times were longer than 1 s (excluding displays that drivers responded to accidentally by pressing the same button again, or in order to "get rid of" the Arrows Task). According to this procedure, 92.5% of all presented Arrows Task displays ($N = 9201$) were included in the analysis of the proportion of display glance time. Again, the data of four drivers (three drivers in the LDW group, and one driver in the NA group) could not be analysed due to insufficient tracking quality. Due to the removal of extreme values for each individual driver that lay more than 2.5 *SD* above or below the mean (procedure explained in section 4.3.2), again 1.3% of the data was excluded from analysis. Thus, the calculation of the mean proportion of display glance time was finally based on a total of 9081 Arrows Task displays.

For the analysis of the mean proportion of display glance time the same research hypotheses applied as for the mean duration of drivers' single glances to the Arrows Task display. It was assumed that drivers' increasing reliance on the lane keeping assistance systems would be reflected by drivers attaching less importance to inspection of the road scene during performance of the Arrows Task, resulting in a higher proportion of display glance time.

At first it was examined whether the two subgroups of HC drivers who had been found to differ substantially in their single display glance durations also differed in their proportions of display glance time. Because the number of drivers in the two HC subgroups was small, it was decided to base the decision of whether they should be treated separately in the following analyses on the magnitude of their difference (effect size) rather than on a significant test result. It was decided to treat the two HC subgroups separately from one another when the effect of their difference for at least one of the three levels of the within-subjects factor had a medium size ($r \geq |.3|$). The analysis revealed that drivers in the 'HC long' subgroup who had been found to glance significantly longer away from the road scene than drivers in the 'HC short' subgroup, also glanced for a higher proportion of time to the in-vehicle display during performance of the Arrows Task than drivers in the 'HC short' subgroup. The effect of their difference was largest in Session 1, $t(13) = 1.49$, $p = .08$ (one-tailed), $r = .38$, and less pronounced for Session 2, both during normal driving periods, $t(13) = .71$, $p = .25$ (one-tailed), $r = .19$, as well as during critical driving situations, $t(13) = .83$, $p = .21$ (one-tailed), $r = .22$. Because the two subgroups of HC drivers differed considerably in their mean proportion of display glance time in Session 1 ($r > .3$), they were treated separately in the following analyses.

Figure 32 depicts the mean proportion of display glance time for the two subgroups of HC drivers and for LDW and NA drivers over the levels of the repeated-measures factor. Group means and standard deviations can be found in Table G1 in Appendix G.

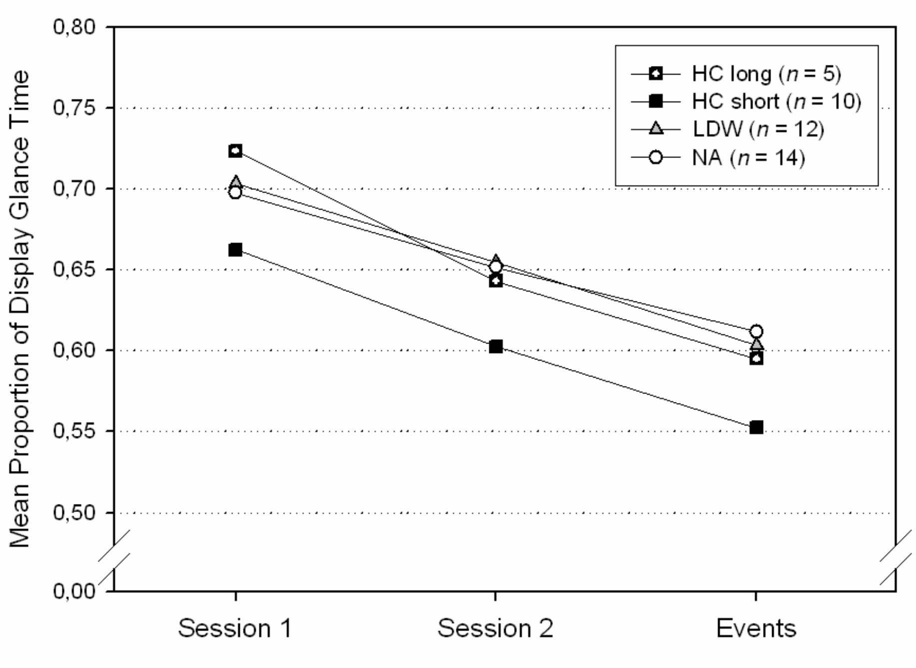


Figure 32. Mean proportion of time that drivers glanced to the in-vehicle display during Arrows Task performance for the two subgroups of HC drivers and for LDW and NA drivers. Mean proportions are aggregated over approximately 94, 92, and 36 Arrows Task displays for each driver in Session 1, Session 2, and in Events, respectively. The linear decrease in the mean proportion of display glance time from Session 1 over normal driving periods in Session 2 to critical driving situations in Session 2 (Events) is highly significant ($r = .41$).

It was assumed that drivers' experience of functional system limits in critical driving situations in Session 2 would lead to a decline in their reliance on the lane keeping assistance systems (reflected by a decline in the mean proportion of display glance time), not only at the very instance of encountering critical situations (Events), but also during normal driving periods in Session 2 as a consequence of drivers' comprehension that they could not rely under any circumstances on the lane keeping assistance systems' support. Thus, it was predicted that there would be a linear decrease in drivers' mean proportion of display glance time from Session 1 (where no critical driving situations occurred) to normal driving periods in Session 2, and from normal driving situations in Session 2 to critical driving situations. Again, this pattern was clearly supported by the data, as can be seen in Figure 32.

A planned contrast yielded evidence that the effect of the linear decrease in drivers' mean proportion of display glance time from Session 1 over normal driving periods in Session 2 to critical driving situations (Events) was even larger than the effect found for the decrease in mean single display glance durations, $t(37) = 9.25$, $p < .001$ (one-tailed), $r = .41$. Thus, drivers' decline in reliance in response to their experience of critical driving situations (partly representing functional system limits) was more strongly reflected by the proportional increase in time they glanced to the road scene during performance of the Arrows Task than by a decrease in the duration of their single glances away from the road scene. Pairwise contrast revealed that the effect was slightly larger for the decline in the mean proportion of display glance time from Session 1 to normal driving periods in Session 2, $t(37) = 5.44$, $p < .001$, $r = .29$, than for the decline from normal driving periods in Session 2 to critical driving situations, $t(37) = 6.63$, $p < .001$, $r = .24$. Again, both effects were larger than those found for the decline in drivers' mean duration of single display glances.

A second hypothesis concerned the effect of the level of lane keeping assistance on drivers' reliance. It was assumed that, previously to drivers' experience of functional system limits, drivers are prepared to allocate higher levels of control to a lane keeping assistance system (reflected by their higher preparedness to allocate their visual attention away from the road scene) the more completely it automates the lateral control task. Thus, in Session 1, HC drivers were assumed to look for a higher proportion of time away from the road scene to the Arrows Task display, whereas the opposite was expected for drivers in the No Assistance group. Although LDW drivers were basically not relieved from lateral control, they were assumed to spend less time than NA drivers for inspection of the road scene during Arrows Task performance in Session 1 given that the LDW system would warn them of a near-by crossing of the lane markings.

As Figure 32 indicates, only drivers in the 'HC long' subgroup glanced for a higher proportion of time to the Arrows Task display in Session 1 than LDW and NA drivers. The effect was however negligible, $t(37) = .55$, $p = .29$ (one-tailed), $r = .09$. On the contrary, drivers in the 'HC short' subgroup glanced for a smaller proportion of time to the display than LDW and NA drivers. This effect was larger but nevertheless rather small, $F(1, 37) = 1.43$, $p = .24$ (two-tailed), $r = .20$. LDW drivers and NA drivers did not differ in their mean proportion of display glance time in Session 1. The results for drivers' mean proportion of display glance time in Session 1 thus replicate the direction of the effects observed for drivers' mean single display glance duration; however, the magnitude of the effects was much smaller.

It was further hypothesised that drivers would react differently to the unexpected occurrence of critical driving situations depending on the level of lane keeping assistance. Thus, HC drivers were assumed to be most affected by their experience of functional system limits, and thus should have the smallest proportions of display glance time in Session 2 and in critical driving situations, followed by LDW drivers. As Figure 32 shows, only drivers in the 'HC short' subgroup had smaller proportions of display glance time in Session 2 and in critical driving situations than LDW drivers; whereas LDW drivers, NA drivers and drivers in the 'HC long' subgroup did almost not differ in the proportion of time they glanced to the Arrows Task display in Session 2 and in critical driving situations. A planned comparison that contrasted drivers in the 'HC short' subgroup with LDW drivers revealed that drivers in the 'HC short' subgroup did not have considerably smaller mean proportions of display glance time than LDW drivers in Session 2 and in critical driving situations, $t(37) = 1.28$, $p = .10$ (one-tailed), $r = .20$.

A fourth hypothesis concerned the magnitude of the decline in drivers' mean proportion of display glance time from normal driving periods in Session 1 to normal driving periods in Session 2 as a function of the level of lane keeping assistance. It was assumed that the decline should be most pronounced for HC drivers, and least pronounced for NA drivers. As shown in Figure 32, LDW drivers and NA drivers did not differ in their decline from Session 1 to Session 2. However, according to the predictions, drivers in the 'HC long' subgroup showed a somewhat larger decline in their mean proportion of display glance time than LDW drivers. This effect was found to be of a medium size, $t(5.41) = .90$[11], $p = .20$ (one-tailed), $r = .27$.

4.3.5 Summary and discussion

Two new measures of drivers' visual attention allocation strategies were used to study drivers' reliance on the lane keeping assistance systems. Previous empirical evidence suggests that reliance on automation is not a stable behaviour, but that it can be viewed as a dynamic process of operators allocating varying levels of control to automation based on its actual performance, situational demands, and the subjective state of the operator.

The two measures used in this study were assumed to be sensitive to dynamic changes in drivers' reliance. High reliance on automation is associated with drivers' less active engagement in the automation-controlled processes, and with the delegation of responsibility towards the automation for maintaining an appropriate task performance. Because lane keeping requires the driver's continuous visual attention to the road, high reliance on the lane keeping assistance systems was assumed to be reflected by drivers' increased preparedness to allocate their visual attention away from the road scene during concurrent performance of a visual demanding secondary task. This increased preparedness of drivers to allocate their visual attention away

---

[11] *t* and *df* adjusted for unequal variances among groups.

from the road scene should be manifested in a longer mean duration of single glances to the secondary task in-vehicle display, as well as in a higher proportion of glance time to the display during performance of the secondary task.

The most striking result was the large difference in drivers' reliance on the HC system. Drivers were predicted to rely to a higher extent on the HC system than on the LDW system before their encountering of critical driving situations, because it was assumed that the reduced need to actively intervene in lateral control would allow HC drivers to concentrate to a higher extent on the secondary task while the HC system would protect them from unacceptable deteriorations in primary task performance.

Obviously however, only part of the HC drivers showed this predicted pattern in behaviour, but this to a surprisingly large extent. Thus, five drivers in the HC group were found to rely to a very high degree on the HC system, as evidenced by their substantially longer single glances away from the road scene compared to both LDW and NA drivers in Session 1 (r = .56). On the contrary, a second part of the HC drivers (n = 10) were found to even refuse to rely on the HC system at all, as evidenced by their substantially shorter single glances away from the road scene compared to LDW and NA drivers in Session 1 (r = .34). The difference in single display glance durations between the two subgroups of HC drivers was large (r = .65) and highly significant, also in Session 2 (r = .52) and in critical driving situations (r = .48).

It was assumed that drivers' experience of critical driving situations that occurred unexpectedly for them in Session 2 would result in a decrease of their preparedness to allocate their visual attention away from the road scene, not only at the very encountering of these situations, but also during 'normal' driving periods in between of those critical situations as a response to drivers' anticipation that they might be required to react suddenly in order to avoid a potential collision with other traffic participants. This linear decrease from Session 1 (where no critical driving situations occurred) over normal driving periods in Session 2 to critical driving situations was evident for both measures of drivers' reliance and proved to be highly significant. It became also apparent that the effect was larger for the decrease in drivers' mean proportion of display glance time (r = .41) than for the decrease in their mean single display glance duration (r = .28). Thus, drivers reacted to the increasing driving task demands foremost by increasing their glance times to the road *between* their glances to the secondary task display rather than by reducing the duration of their glances away from the road scene themselves.

It was further predicted that drivers react differently to the unexpected occurrence of critical driving situations in Session 2 that partly represented functional limits of the HC and LDW system. Thus, it was assumed that HC drivers would be most affected by their experience of functional system limits because of the HC system's greater intervention in lateral control and the larger negative consequences for performance and safety when its support was inadequate. LDW drivers were assumed to be less affected by their experience of functional system limits because they were still manually engaged in lateral control and could "simply" ignore the LDW system's warnings when they were inadequate under specific circumstances. Drivers in the control group were assumed to be least affected by the occurrence of critical driving situations.

It turned out that only those HC drivers who relied to a large extent on the HC system in Session 1 as measured by the mean duration of their single display glances (‘HC long’ subgroup) showed a larger decline in reliance than LDW drivers in response to their experience of functional system limits, as evidenced by their larger decrease in mean single display glance durations from Session 1 to Session 2 (r = .50). On the contrary, those HC drivers who were substantially less prepared to allocate their attention away from the road scene (‘HC short’ subgroup) than LDW drivers in Session 1 did not differ significantly from LDW drivers in their decline in mean single display glance duration (r = .17). This latter finding is indeed not surprising because the ‘HC short’ subgroup generally appeared to refuse to rely on the HC system. It appears reasonable that they did not show a larger decline in reliance than LDW drivers. Still, their decline in mean single display glance duration from Session 1 to Session 2 was more pronounced than that of the LDW drivers, presumably because of the HC system’s higher degree of intervention in lateral control even when they did not rely on the system. The effects found for the second measure of drivers’ reliance – the mean proportion of display glance time – went in the same direction but were much smaller in magnitude.

It was also found that LDW drivers did not differ from NA drivers both in the mean duration of their single glances away from the road scene as well as in the proportion of time they glanced to the in-vehicle display during performance of the secondary task. Since the data of the NA drivers served as a baseline, this finding suggests that LDW drivers did in fact not rely on the LDW system.

It turned also out that the mean proportion of display glance time was a less sensitive measure than the mean duration of single display glances for the effects of the level of lane keeping assistance on drivers’ reliance. Although the HC subgroup with the long single display glance durations also looked for a higher proportion of time to the in-vehicle display during performance of the Arrows Task compared to the HC subgroup with the short single display glance durations, this effect was of much smaller magnitude than the effect found for the two HC subgroups’ difference in mean single display glance duration. In fact, the results of the mean proportion of display glance time rather indicate that especially drivers in the ‘HC long’ subgroup seemed to partly compensate for their extensive glances away from the road scene by looking for longer times to the road scene afterwards. Thus, it seems that drivers needed more time to update their mental representation of the actual traffic situations after they had been looking to the in-vehicle display for longer periods of time.

4.3.6 Drivers’ compensatory reduction of primary task demands: Analysis of drivers’ speed

There is empirical evidence that drivers compensate for the increasing visual demands imposed by a secondary task while driving by a reduction of speed (e.g., Engström et al., 2005). This has been interpreted as an active attempt of drivers to lower primary driving task demands in order to protect driving performance under the dual task conditions. In order to rule out that especially the two HC subgroups' difference in reliance on the HC system was attributable to their different compensation strategies, drivers' speed behaviour was analysed. Particularly, it was investigated whether the longer eyes-off-the-road times of drivers in the 'HC long' subgroup could be attributed to a greater reduction of their speed during concurrent performance of the Arrows Task.

This was apparently not the case. Figure 33 shows drivers' mean speed during driving periods with concurrent Arrows Task performance in Session 1 and Session 2 for the two subgroups of HC drivers, and for LDW and NA drivers[12]. Group means and standard deviations can be found in Table H1 in Appendix H.

While the two subgroups of HC drivers did not differ in their mean speed in Session 1, they were found to differ remarkably in Session 2. However, the direction of their difference in mean speed speaks against the assumption that the longer eyes-off-the-road times of drivers in the 'HC long' subgroup (who were found to rely to a high extent on the HC system) can be attributed to their greater reduction of speed as indication for a greater reduction of primary task demands. At least in Session 2, drivers in the 'HC long' subgroup drove faster than drivers in the 'HC short' subgroup (who had the shortest single glances away from the road scene).

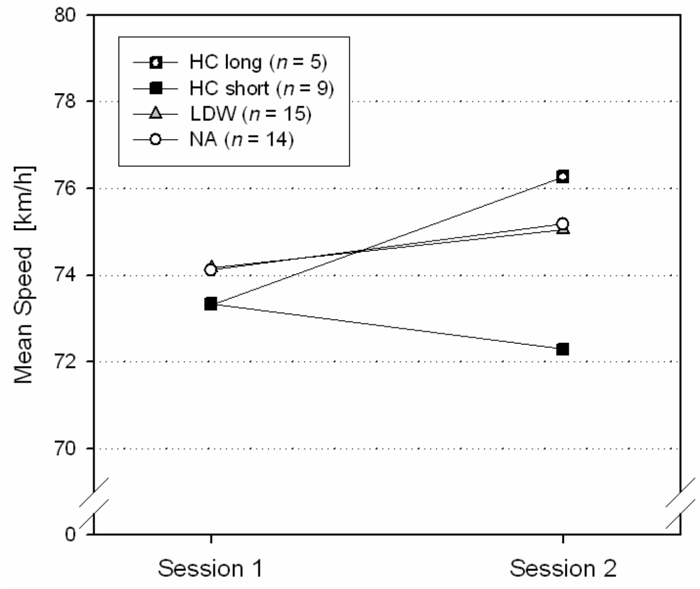


Figure 33. Drivers' mean speed during driving periods with concurrent Arrows Task performance in Session 1 and Session 2 as a function of the level of lane keeping assistance, separately for the two subgroups of HC drivers that were found to differ significantly in their reliance on the HC system.

---

[12] One driver in the 'HC short' subgroup and one NA driver were excluded from the analysis due to extreme values that lay more than 2.5 *SD* above the group mean. Their values were also identified as outliers in driving periods without concurrent Arrows Task performance.

A 4 (level of lane keeping assistance: HC long, HC short, LDW and NA) x 2 (session) repeated-measures ANOVA with the level of lane keeping assistance as between-subjects factor and session as within-subjects factor revealed a significant effect of session, $F(1, 39) = 4.85$, $p = .03$, $\eta^2_G = .01$. The interaction between level of lane keeping assistance and session only approached statistical significance, $F(3, 39) = 2.69$, $p = .06$, $\eta^2_G = .01$. There was no significant main effect of the level of lane keeping assistance, $F(3, 39) = .31$, $p = .81$, $\eta^2_G = .02$.

4.4 Factors influencing Drivers' Reliance on the Lane keeping Assistance Systems

4.4.1 Drivers' strategic investment of mental effort

Motivational processes, i.e. drivers' strategic decision about the investment of mental effort were assumed to play a major role in influencing their reliance on the lane keeping assistance systems. Specifically it was assumed that higher levels of reliance on a lane keeping assistance system are (at least partially) the result of drivers' motivation to limit the expenditure of mental effort in the driving task while still maintaining an acceptable driving performance due to the lane keeping assistance systems' support. Thus, it was assumed that high reliance on a lane keeping assistance system is accompanied by drivers' reduced investment of mental effort in the primary driving task.

Drivers' mental effort regulation strategies were assessed through different means. First, they were assessed rather implicitly by referring to the magnitude of latent performance decrements associated with the expenditure of mental effort. The central hypothesis in this regard was that drivers who rely to a high degree on a lane keeping assistance system should have invested lower mental effort in the primary driving task, and this conservation of mental effort should in turn become apparent in lower costs in the form of latent performance decrements. Latent performance decrements were assumed to occur in two forms: (a) as compensatory costs, which are the physiological and affective side effects of humans' attempt to protect task performance under high workload conditions, and (b) as subsidiary task failures, which refer to the selective impairment of lower priority task components. The magnitude of compensatory costs was assessed by subjective measures of drivers' mental effort, workload and frustration. The magnitude of subsidiary task failures was assessed by objective measures of drivers' secondary task performance. Accordingly, it was assumed that high reliance on the lane keeping assistance systems is associated with drivers' superior performance in the secondary task, as measured by response accuracy and response times.

Second, drivers' strategies for mental effort investment were assessed subjectively, e.g. by asking drivers about how they prioritised the driving task over the secondary task.

The results obtained for these different measures of drivers' mental effort regulation strategies will be reviewed consecutively in the following sections.

4.4.1.1 Subsidiary task failures: Drivers' secondary task performance

Two measures of drivers' performance in the Arrows Task were analysed: first, drivers' response times to the Arrows Task displays; and second, response accuracy as indicated by the percentages of drivers' correct, false, and missing responses to the Arrows Task.

Drivers' mean response time to the Arrows Task displays

Drivers' response times were first aggregated over six consecutive Arrows Task sections consisting of six Arrows Tasks each. The first three Arrows Task sections belonged to Session 1, and the second three Arrows Task sections belonged to Session 2. The Arrows Tasks where critical situations occurred were treated separately and constituted an extra category. For the six Arrows Task sections and the extra category, extreme response times that lay 2.5 *SD* above or below the mean for each individual driver were removed from the data (corresponding to 1.1% of the data). The three Arrows Task sections of each session were then aggregated to obtain one value for each driver for Session 1 and for Session 2 (excluding critical situations).

It was hypothesised that drivers' increasing reliance on a lane keeping assistance system would result (at least partially) from their adoption of an effort conservation strategy whereby they were able to limit the expenditure of mental effort in the driving task. It was predicted that drivers rely to a higher extent on the HC system than on the LDW system in Session 1 before they experienced functional system limits. The results on drivers' reliance however indicated that only one subgroup of the HC drivers ('HC long' subgroup) showed this predicted higher level of reliance on the HC system, whereas the other subgroup of HC drivers ('HC short' subgroup) did not rely on the HC system at all, and even showed a smaller preparedness to allocate their visual attention away from the road scene than NA drivers (whose attention allocation strategies served as a baseline). If drivers' reliance on the lane keeping assistance systems was in fact mediated by drivers' strategic investment of mental effort, then it would be predicted that drivers in 'HC long' subgroup had invested the least mental effort in the driving task in Session 1. Correspondingly, drivers in the 'HC long' subgroup should have suffered to the least extent from latent performance decrements, and should therefore show the best secondary task performance in terms of the shortest response times to the Arrows Task. On the contrary, drivers in the 'HC short' subgroup could be assumed to have invested the highest level of mental effort in the driving task in Session 1, which in turn should become apparent in the largest performance decrements in the secondary task (corresponding to the longest response times). LDW drivers were assumed to have invested less effort in the driving task than NA drivers in Session 1, and should therefore show a better secondary task performance.

The data in Figure 34 does not support these predictions.

Depicted in Figure 34 are the mean response times for the three levels of lane keeping assistance, separately for the two subgroups of HC drivers that were found to differ substantially in their reliance on the HC system (for group means and standard deviations refer to Table I1 in Appendix I).

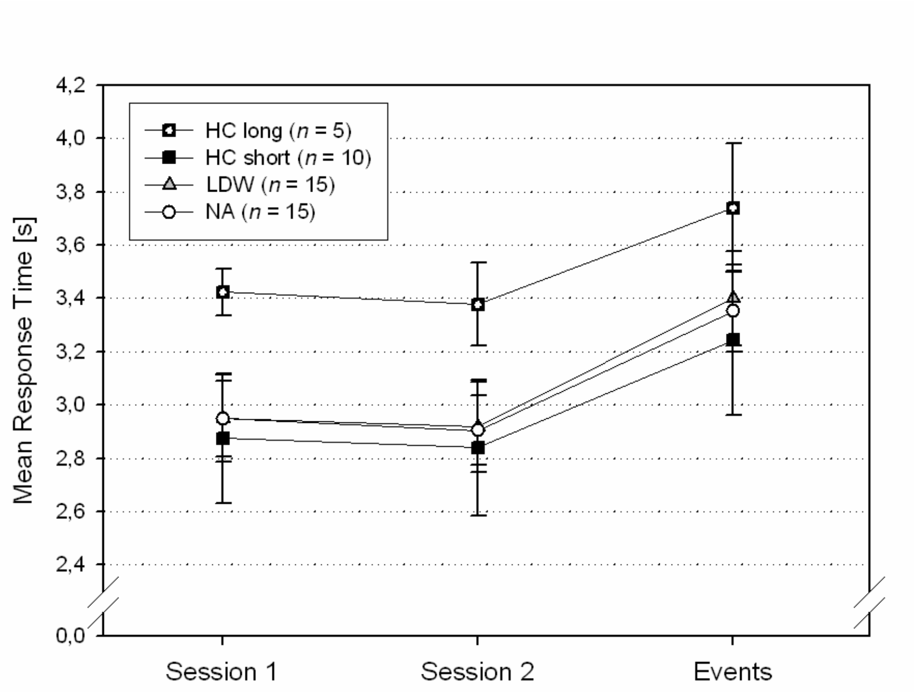


Figure 34. Drivers' mean response times to the Arrows Task displays during normal driving periods in Session 1 and in Session 2, and during critical driving situations (Events) as a function of the level of lane keeping assistance. Mean response times are aggregated over approximately 108, 94, and 51 Arrows Task displays for each driver in Session 1, Session 2, and in Events, respectively. Mean response times are displayed separately for the two subgroups of HC drivers that were found to differ significantly in their reliance on the HC system. The two HC subgroups did not differ significantly in their response times. Drivers' mean response time was significantly longer in critical driving situations than in Session 1 and Session 2. Error bars depict standard errors.

Indeed, the two HC subgroups of drivers were found to differ considerably ($r \geq |.3|$, corresponding to a medium effect size) in their mean response time to the Arrows Task. Since there was no interaction between the level of lane keeping assistance and the within-subjects factor session-event as can be seen in Figure 34, the two subgroups of HC drivers were contrasted over all three levels of the within-subject factor, which yielded $F(1, 41) = 2.45$, $p = .12$ (two-tailed), $r = .34$. Although the two subgroups of HC drivers were found to differ in their response times to the Arrows Task, the difference was actually in the opposite direction. Contrary to what would have been assumed, drivers in the 'HC long' subgroup (who relied to a high extent on the HC system) did not have the shortest response times, but did instead have substantially longer response times (about 3.4 s in Session 1 and 2) than all other groups of drivers. On the other hand, drivers in the 'HC short' subgroup (who did not rely at all on the HC system) had even slightly shorter response times (about 2.85 s in Session 1 and Session 2) than both LDW and NA drivers.

It was further predicted that drivers' response times to the Arrows Task increase from Session 1 to Session 2 as a function of their experience of critical driving situations (that partly represented functional limits of the HC and LDW system). Drivers' perception that they might be required to react suddenly in order to avoid a potential collision together with their perception that they could not rely under any circumstances on the lane keeping assistance systems was hypothesised to lead to a shift in their mental effort regulation strategies. Specifically, it was hypothesised that drivers invest more effort in the driving task in Session 2 in order to adapt to the increasing task demands. This shift to more resource-intensive performance strategies was assumed to result in greater costs occurring in the form of latent performance decrements under dual-task conditions.

Thus, the linear decline in drivers' preparedness to allocate their visual attention away from the road scene from Session 1 to normal driving periods in Session 2 over to critical driving situations should be mirrored by a *linear increase* in drivers' response times. Interestingly, although drivers showed the predicted pattern in their reliance on the lane keeping assistance systems, they did not show the predicted pattern in their response times to the Arrows Task. It was found that drivers' response times did not increase from Session 1 to Session 2; however, response times increased markedly during critical driving situations, as Figure 34 shows.

HC drivers were expected to show the worst secondary task performance (corresponding to the longest response times) in Session 2 and in critical driving situations because they were assumed to be most affected by their experience of functional system limits. On the other hand, LDW and NA drivers were predicted to show a similar better performance in the Arrows Task than HC drivers in Session 2 and in critical situations. As Figure 34 shows, there was however no evidence for an interaction between the level of lane keeping assistance and the session-event factor. Instead, the differences between the levels of lane keeping assistance appeared to remain stable over the levels of the session-event factor.

In order to test whether the increase in drivers' response times from Session 2 to critical driving situations (Events) was significant, an ANOVA was conducted with the level of lane keeping assistance (HC long, HC short, LDW, and NA) as between-subjects factor and session-event as within-subjects factor.

Indeed, the main effect of session-event was found to be highly significant, $F(1, 41) = 35.86$, $p < .001$, $\eta^2_G = .07$ (degrees of freedom adjusted with lower bound-ε for non-sphericity). Post-hoc comparisons revealed that drivers' response times in critical driving situations (Events) were significantly longer than in Session 1, $F(1, 41) = 32.09$, $p < .001$, $r = .26$, and also than in Session 2, $F(1, 41) = 78.94$, $p < .001$, $r = .28$.

As supposed, there was no interaction between session-event and the level of lane keeping assistance, $F(3, 41) = .15$, $p = .93$, $\eta^2_G = .00$ (*df* adjusted with lower bound-ε for non-sphericity). As supposed by the non-significant difference in response times of the two HC subgroups, also the main effect of the level of lane keeping assistance was not significant, $F(3, 41) = .88$, $p = .46$, $\eta^2_G = .05$.

Response Accuracy

Response accuracy was the second measure of drivers' secondary task performance used to evaluate the magnitude of subsidiary task failures as a form of latent performance decrement under high workload conditions. Response accuracy was measured by the percentages of drivers' correct, false, and missing responses to the Arrows Task. Those percentages were first calculated for each single Arrows Task, and then averaged over all Arrows Tasks in Session 1 ($n = 19$), over all Arrows Tasks that occurred during normal driving periods in Session 2 ($n = 16$), and over all Arrows Tasks that occurred during critical driving situations ($n = 9$) in order to form the three levels of the session-event factor.

The same hypotheses as for drivers' response time to the Arrows Task also applied for the analysis of drivers' response accuracy. Specifically, it was assumed that higher levels of reliance on the lane keeping assistance systems in Session 1 would be associated with a lower magnitude of subsidiary task failures, and hence a better secondary task performance (reflected by a higher response accuracy) in Session 1.

As the analysis of drivers' reliance on the lane keeping assistance systems revealed, HC drivers could be split in two subgroups who differed considerably in their reliance on the HC system. Exploratory data analysis revealed however that the two HC subgroups did *not differ* in their response accuracy to the Arrows Task. Because of this, they were not treated separately in the analysis.

Figure 35 to Figure 37 depict the mean percentages of drivers' correct, false, and missing responses to the Arrows Task over the three levels of the session-event factor as a function of the level of lane keeping assistance (for group means and standard deviations refer to Tables I2 to I4 in Appendix I). In general, drivers performed quite well in the Arrows Task with about 90% correct responses in Session 1 and Session 2. In critical driving situations, response accuracy markedly decreased to an average of 72% correct responses.

There was a general trend that HC drivers showed the worst secondary task performance in terms of response accuracy, whereas NA drivers showed the best performance. This pattern of results had been predicted however only for Session 2 and for critical driving situations as a result of drivers' experience of functional system limits, but not for Session 1. In Session 1, the lower response accuracy of the HC drivers would have been only predicted for the 'HC short' subgroup that did not rely on the HC system at all. On the contrary, the 'HC long' subgroup would have been expected to show the best response accuracy in Session 1. Also, a reverse pattern was predicted for LDW and NA drivers in Session 1. In fact, NA drivers were supposed to show a worse response accuracy than LDW drivers in Session 1.

Paralleling the results of drivers’ response times to the Arrows Task, response accuracy did not decline from Session 1 to Session 2 as a function of drivers’ experience of critical driving situations, against the predictions. However, there was a marked decrease in drivers’ response accuracy in critical driving situations. It seemed that drivers had adopted different performance strategies in critical driving situations. Whereas LDW drivers apparently tried to continue to perform the Arrows Task during critical driving situations (while accepting a higher percentage of false responses), HC drivers tended to skip the Arrows Task, which in turn resulted in a higher percentage of missing responses during critical driving situations.

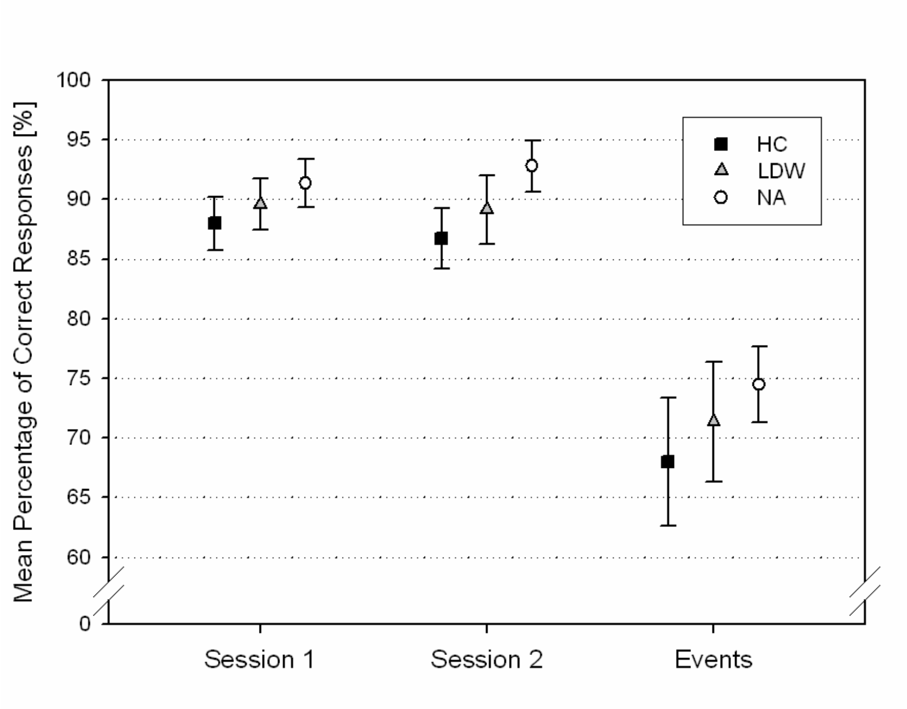


Figure 35. Mean percentage of drivers' correct responses to the Arrows Task in normal driving periods in Session 1 and Session 2 and in critical driving situations (Events) as a function of the level of lane keeping assistance. Error bars depict standard errors.

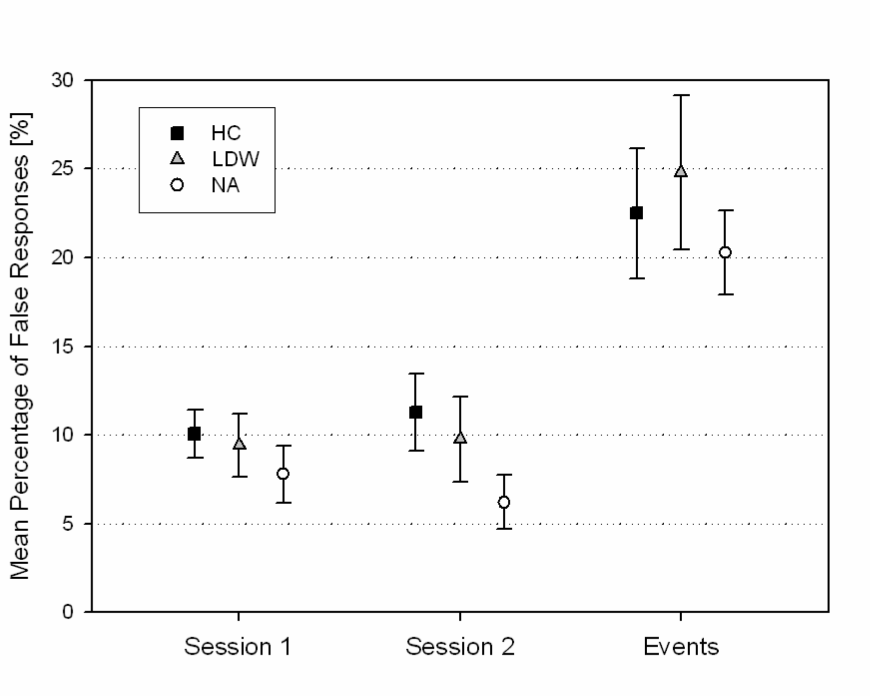


Figure 36. Mean percentage of drivers' false responses to the Arrows Task in normal driving periods in Session 1 and Session 2 and in critical driving situations (Events) as a function of the level of lane keeping assistance. Error bars depict standard errors.

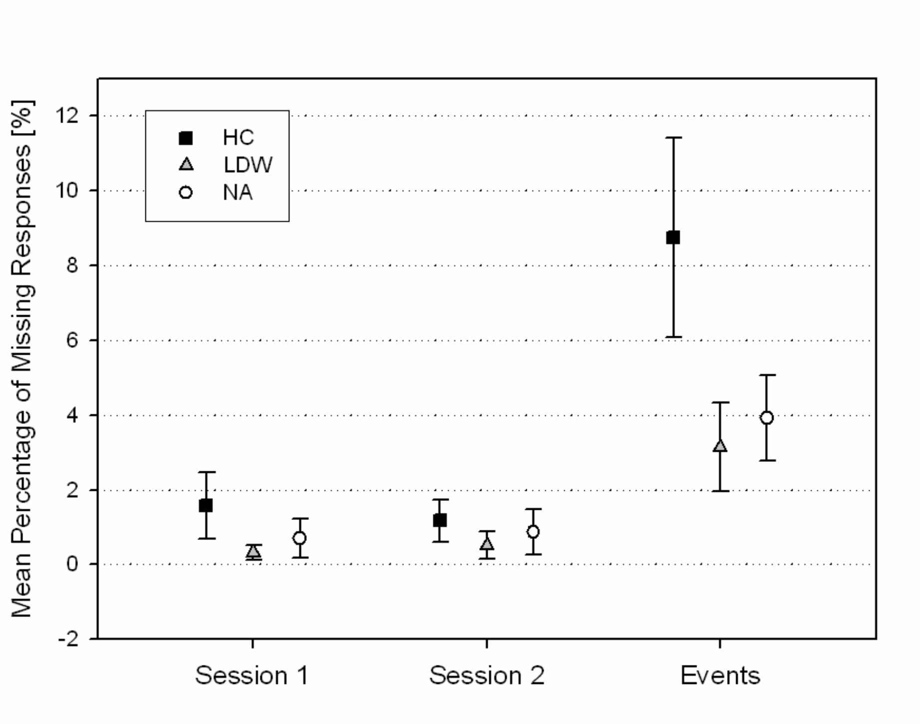


Figure 37. Mean percentage of drivers' missing responses to the Arrows Task in normal driving periods in Session 1 and Session 2 and in critical driving situations (Events) as a function of the level of lane keeping assistance. Error bars depict standard errors.

For statistical analysis, the three measures of drivers' response accuracy (the mean percentages of correct, false, and missing responses to the Arrows Task) were combined into one single index. The index was calculated for each Arrows Task and reflected the proportion of the number of displays answered to the number of displays presented within each Arrows Task, while displays answered correctly obtained a higher weight than displays answered wrong. The index could vary between 0 (corresponding to 100% missing responses) and 1 (corresponding to 100% correct responses) and was calculated as

$$Accuracy\ Index = \frac{n_{answered} + n_{correct}}{n_{answered} + n_{presented}} \tag{2}$$

where $n_{answered}$ is the number of displays that drivers responded to during one Arrows Task (excluding missing responses), $n_{correct}$ is the number of correctly answered displays, and $n_{presented}$ is the total number of displays that were presented during this Arrows Task.

Figure 38 shows the mean accuracy index for the three groups of drivers for Session 1, Session 2, and for critical driving situations (Events). For group means and standard deviations refer to Table I5 in Appendix I.

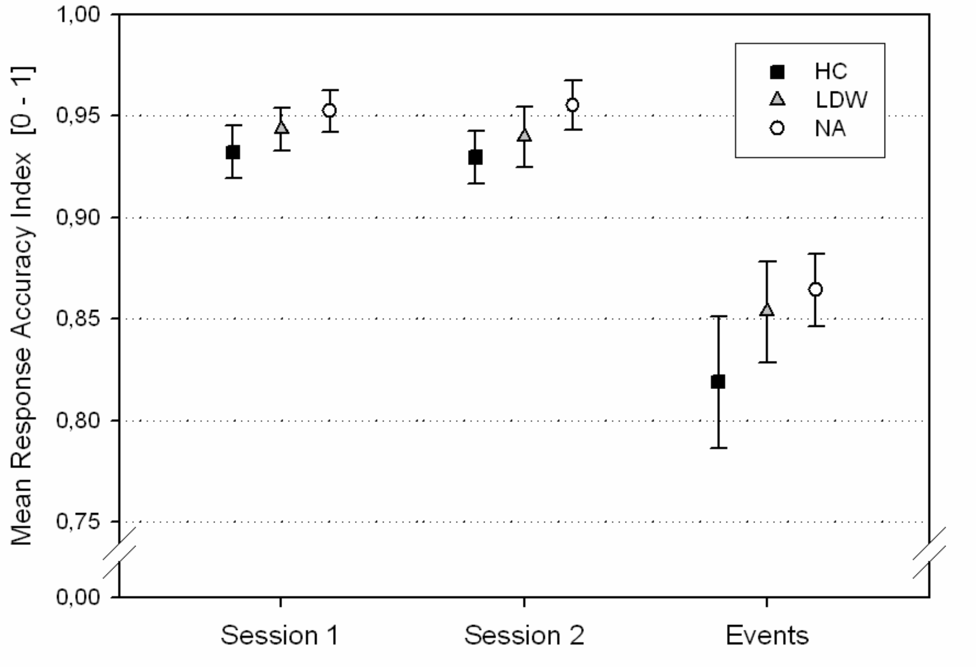


Figure 38. Mean response accuracy index as a combined measure of drivers' mean percentages of correct, false, and missing responses to the Arrows Task during normal driving periods in Session 1 and 2 and during critical driving situations (Events) as a function of the level of lane keeping assistance. The decline in drivers' response accuracy in critical driving situations is highly significant. Error bars depict standard errors.

An ANOVA of the log-transformed[13] response accuracy index was conducted with the level of lane keeping assistance as between-subjects factor and session-event as within-subjects factor. One LDW driver was considered as outlier (laying more than 2.5 *SD* above the group means) and was therefore excluded from the analysis.

The ANOVA revealed that the three groups of drivers did not differ significantly in their response accuracy, $F(2, 41) = 1.34$, $p = .27$, $\eta^2_G = .05$. Also, the interaction between the level of lane keeping assistance and session-event, i.e. the slightly larger decline in response accuracy of the HC drivers in critical driving situations (Events) was not statistically significant, $F(4, 82) = 1.35$, $p = .26$, $\eta^2_G = .01$. However, there was a large main effect of session-event, $F(2, 82) = 91.10$, $p < .001$, $\eta^2_G = .31$. Post-hoc comparisons confirmed that drivers' response accuracy in critical driving situations was considerably worse compared to normal driving periods in Session 1, $F(1, 41) = 122.13$, $p < .001$, $\eta^2_G = .35$, and in Session 2, $F(1, 41) = 167.29$, $p < .001$, $\eta^2_G = .34$.

Summary and discussion

One major assumption in this study was that drivers' reliance on lane keeping assistance systems is influenced by motivational processes, i.e. by their strategic investment of mental effort. It was assumed that driver assistance systems offer the potential for drivers to invest less mental effort in the driving task while primary task performance can still be kept at a high level due to the systems' support. It was hypothesised that high reliance on a lane keeping assistance system is (at least partially) the result of drivers' motivation to limit their expenditure of mental effort in the driving task.

Drivers' secondary task performance was used as an indirect measure of drivers' mental effort regulation strategies. The hypothesis was that high reliance on a lane keeping assistance system is associated with drivers' reduced investment of mental effort in the driving task. As a consequence, high reliance on a lane keeping assistance system should be accompanied by smaller performance decrements in lower priority tasks. In this case, the secondary task (Arrows Task) was considered as a task of lower priority.

The results of drivers' secondary task performance let suggest that drivers did not regard the Arrows task as a task of low priority. This is indicated by the fact that drivers' secondary task performance did not deteriorate from Session 1 (where no critical driving situations occurred) to 'normal' driving periods in Session 2, although drivers encountered eight unexpected critical driving situations in Session 2. This finding is particularly interesting because drivers *did* respond to the unexpected occurrence of critical driving situations by changing their visual attention allocation strategies. Thus, as measures of drivers' eye glance behaviour demonstrate, there was a marked decline in drivers' preparedness to allocate their visual attention away from

---

[13] The distribution of the response accuracy index was highly skewed to the right. For statistical analyses the Index measure was transformed with Index′ = $\log \frac{Index}{1-Index}$.

the road scene from Session 1 to Session 2; however, drivers' secondary task performance did not deteriorate as a consequence. Two possible explanations may account for this finding. First, drivers may have switched to more effective performance strategies in Session 2 in response to the increasing driving task demands; and second, drivers may have mobilised some extra effort in order to protect secondary task performance in Session 2 despite the increasing primary task demands. The two explanations are not mutually exclusive. It is indeed plausible that drivers did not invest their maximum of available resources at any time during the drive. Drivers might have been capable to perform the Arrows Task concurrently while driving in Session 1 without expending their maximum effort budget while reserving some resources for unexpected increases in task demands. Theoretical frameworks assume that mental effort investment is based on cost-benefit decisions about the use of effort and the relative value of different goals (Hockey, 1997). As long as drivers were capable to maintain dual-task performance at a high standard in Session 1 it seems irrational to assume that they tried to even improve performance through maximising their effort investment because operating at higher levels of effort for any length of time is known to be uncomfortable and avoided whenever possible.

As driving task demands increased in Session 2 through drivers' unexpected encountering of critical driving situations, drivers apparently switched to more effective time-sharing strategies that enabled them to perform the Arrows Task as fast and as accurate as in Session 1 despite that they allocated less visual attention to it. The fact that drivers tried to protect secondary task performance despite the increase in driving task demands in Session 2 argues for the interpretation that they attached considerable importance to the Arrows Task and did not regard it as task of low priority. It can be assumed however that this maintenance of high performance standards in the Arrows Task in Session 2 was only possible at the expense of an increase in compensatory costs, i.e. an increase in drivers' mental workload and frustration. Subjective measures of compensatory costs and drivers' mental effort regulation strategies are going to provide an indication of whether this assumption is valid.

Drivers' shift in mental effort regulation strategies in Session 2 is also substantiated by results in measures of drivers' lane keeping performance. It was found that, in sections without concurrent Arrows Task performance, drivers' lane keeping performance significantly deteriorated from Session 1 to Session 2. It seems that drivers compensated for the increasing effort they had to invest in order to protect secondary task performance by lowering their performance standards in periods without Arrows Task that were less vulnerable to deteriorations in driving task performance.

The marked decline in drivers' Arrows Task performance in critical driving situations indicates that drivers were not able to continue to protect secondary task performance when driving task demands increased even further. Drivers reacted by a marked shift in task priorities in face of the increased safety costs that were associated with their inadequate driving performance in critical driving situations. For HC drivers, this shift in priorities was most pronounced, as predicted based on the larger intervention of the HC system in lateral control. Thus, in critical driving situations, HC drivers were found to stop continuing performance of the Arrows Task to a larger extent compared to LDW and NA drivers, indicated by a higher percentage of missing responses.

A surprising finding was that drivers who were found to rely to a high extent on the HC system ('HC long' subgroup) took considerably longer to respond to the Arrows Task displays than drivers who were found to not rely on the HC system at all ('HC short' subgroup), as well as than LDW and NA drivers. Indeed, it was hypothesised that high reliance on a lane keeping assistance system would be associated with better secondary task performance based on the assumption that high reliance is associated with drivers investing less mental effort in the driving task, which in turn should leave them more resources available for performance of the secondary task. Although the differences in response times were not statistically significant, this finding is hard to explain.

It was assumed that drivers would benefit from less frequent, but longer glances away from the road scene to the Arrows Task display in terms of response time because this strategy was supposed to result in lower costs (time) for visual reorientation compared to an intermittent glance strategy (consisting of subsequent glances away and back to the road scene). However, as already shown in section 4.3.4, the longer glances away from the road scene of drivers in the 'HC long' subgroup were not accompanied by a similar decline in glance time to the road, as demonstrated by the results for drivers' mean proportion of display glance time. Thus, one reason for the prolonged response times of drivers in the 'HC long' subgroup may lay in the relative ineffectiveness of their visual attention allocation strategies, following from a combination of long glances to the Arrows Task display and relatively long glances back to the road scene afterwards.

Another explanation for the 'HC long' subgroup's longer response times may be that those drivers felt less stressed and more relaxed during performance of the secondary task because of the lower driving task demands due to the HC system's assistance. Thus, drivers in the 'HC long' subgroup may not have experienced a "need" for hurrying up with their performance of the secondary task because the HC system helped them to prevent severe deteriorations in lane keeping performance. The relatively long glance times to the road of the 'HC long' subgroup however seems to speak against this interpretation of results.

Alternatively, drivers in the 'HC long' subgroup may have taken "use of" the HC system's support and may have tried to perform particularly well in the Arrows Task compared to LDW and NA drivers who didn't have this possibility to the same extent, as well as compared to the 'HC short' subgroup who didn't allocate control to the HC system. This higher motivation of drivers in the 'HC long' subgroup to perform well in the Arrows Task should be reflected accordingly by subjective measures, which will be reviewed in the next chapters.

Last, another unpredicted finding was the superior performance of the control group (NA) in the Arrows Task in terms of response accuracy in Session 1 compared to both the LDW and HC group, while the LDW group showed again a better Arrows Task performance than the HC group (in terms of response accuracy). According to the hypotheses, an opposite pattern of results had been predicted for Session 1. Also, the two HC subgroups were found to not differ in their response accuracy. The lower response accuracy of drivers in the ‘HC short’ subgroup in Session 1 may be explained by their lacking reliance on the HC system and their lower preparedness to allocate their visual attention away from the road scene, as indicated by their substantially shorter glances to the in-vehicle display during Arrows Task performance. This might have resulted in slightly faster response times, but also in a lower response accuracy. However, this does not explain the worse response accuracy of drivers in the ‘HC long’ subgroup compared to LDW and NA drivers in Session 1 because those HC drivers were found to have considerably longer single display glance durations than both LDW and NA drivers. Aside from that, LDW and NA drivers were found to differ in any of the eye glance behaviour measures that would have helped to explain their differences in response accuracy. At this point however there is no satisfying explanation for this unpredicted pattern of results. However, one has to keep in mind that the differences between the levels of lane keeping assistance (although they went in the opposite direction for Session 1) were quite small and not significant.

4.4.1.2 Compensatory costs: Drivers’ perceived mental workload and frustration

Compensatory costs refer to the physiological and affective side effects of human regulatory behaviour and were considered as a second form of latent performance decrements (besides subsidiary task failures) under the dual-task conditions. The magnitude of compensatory costs was assessed by subjective measures of mental effort, mental workload, and frustration.

It was hypothesised that increasing levels of drivers’ reliance on a lane keeping assistance system would (at least partially) result from drivers’ attempt to limit their expenditure of mental effort in the driving task. Consequently, it was assumed that higher reliance on a lane keeping assistance system would be associated with lower compensatory costs in terms of perceived mental workload and frustration. Taking into account the results on drivers’ reliance on the lane keeping assistance systems, it could be assumed that drivers in the ‘HC long’ subgroup who were found to rely to a high extent on the HC system therefore experienced the lowest level of mental workload and frustration in Session 1, whereas drivers in the ‘HC short’ subgroup who were found to not rely on the HC system at all (and who were even less prepared to allocate their visual attention away from the road scene than NA drivers) experienced the highest level of mental workload and frustration in Session 1.

It was further assumed that drivers experience a higher level of mental workload and frustration in Session 2 compared to Session 1 because of the increase in driving task demands due to the unexpected occurrence of critical driving situations. It was supposed that HC drivers would be most affected by their experience of functional system limits and their need to actively override the system in some driving situations. Thus, HC drivers were assumed to experience the highest level of mental effort in Session 2, whereas LDW and NA drivers were assumed to experience a similar lower level of mental effort than HC drivers in Session 2.

Drivers' perceived mental workload

Drivers' overall mental workload was assessed by the help of several rating scales in the questionnaires, namely by the Rating Scale Mental Effort (RSME), the Raw Task Load Index (RTLX) and the Mental Effort subscale of the NASA-TLX. Those three measures yielded similar results and only the results of the RSME and RTLX will be discussed here in more detail.

Rating Scale Mental Effort (RSME)

Drivers rated their overall mental effort on the RSME at three times during the study: once after the practice session (where they drove with either the HC system, the LDW system, or without lane keeping assistance and *did not* perform the Arrows Task concurrently to driving), once after Session 1 (where they drove with the respective level of lane keeping assistance *and* performed the Arrows Task concurrently to driving) and once after Session 2 (where they drove with the respective level of lane keeping assistance, performed the Arrows Task concurrently to driving, and also encountered eight critical driving situations). The mental effort rating after the practice session was used to investigate the effect of the change from single-task to dual-task conditions on drivers' mental workload.

Figure 39 shows the results for the RSME (for group means and standard deviations refer to Table I6 in Appendix I). As predicted, drivers' perceived mental effort clearly increased from "a little" in the practice session to "rather much" in Session 1 to "considerable" up to "great" in Session 2. This linear increase in mental effort ratings turned out to be large and highly significant, $F(1, 41) = 116.53$, $p < .001$ (one-tailed), $r = .62$. Drivers' perceived mental effort was found to increase to a larger extent due to the change from single-task (driving alone) to dual-task (driving plus performing the Arrows Task) conditions, $F(1, 41) = 56.68$, $p < .001$ (one-tailed), $r = .51$, than due to the increase in task demands resulting from drivers' unexpected encountering of the critical driving situations in Session 2 compared to Session 1, $F(1, 41) = 40.97$, $p < .001$ (one-tailed), $r = .32$.

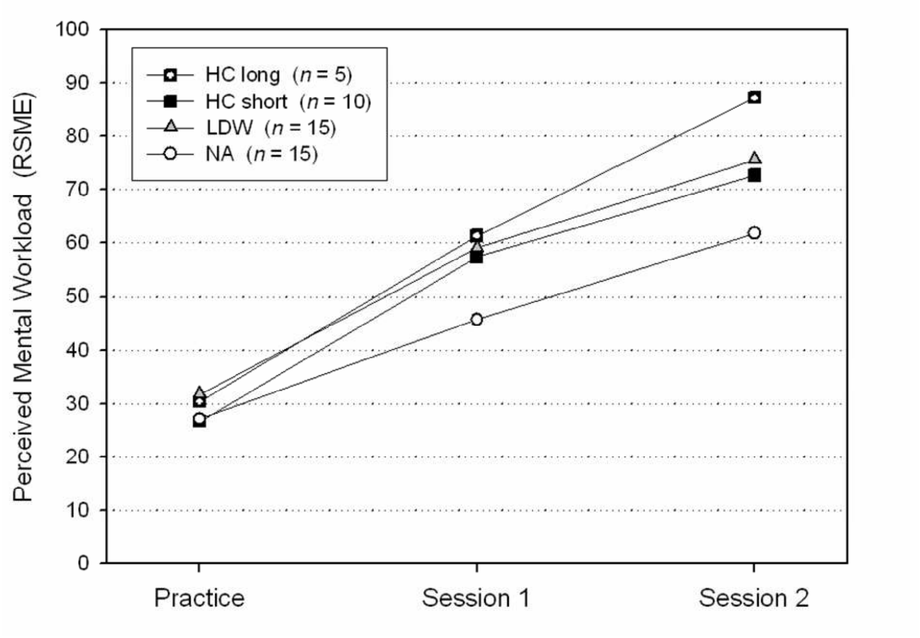


Figure 39. Drivers' ratings of overall mental workload on the Rating Scale Mental Effort (RSME) as a function of the increasing driving task demands from the practice session over Session 1 to Session 2 across the three levels of lane keeping assistance, separately for the two subgroups of HC drivers that were found to differ significantly in their reliance on the HC system. The increase in drivers' mental workload from the practice session over Session 1 to Session 2 is highly significant.

Surprisingly and contrary to the predictions, NA drivers reported a lower level of mental workload than both HC and LDW drivers in Session 1. In fact, NA drivers were assumed to report a higher level of mental workload than LDW drivers in Session 1, which might only have been exceeded by the 'HC short' subgroup's workload ratings. However, interestingly, drivers driving with a lane keeping assistance system reported a similar higher level of mental workload than NA drivers in Session 1. Also, the two subgroups of HC drivers did not differ in their mental workload rating in Session 1, as could have been expected based on their different reliance on the HC system.

In Session 2, NA drivers were assumed to report the lowest level of mental workload, while HC drivers were expected to report the highest level of mental workload, because they should be most affected by their experience of functional system limits. The pattern of results in Session 2 did largely support this prediction, despite that only the 'HC long' subgroup reported a higher level of mental workload than the other groups of drivers. Drivers in the 'HC long' subgroup also reported a higher level of mental workload than drivers in the 'HC short' subgroup in Session 2, $t(13) = 1.19$, $p = .26$ (two-tailed), $r = .31$.

An ANOVA over two levels of the repeated-measures factor (Session 1 and Session 2) and over four levels of the between-subjects factor (HC long, HC short, LDW and NA) revealed that neither the main effect of the level of lane keeping assistance was significant, $F(3, 41) = 1.33$, $p = .28$, $\eta^2_G = .08$, nor was there a significant Level of lane keeping assistance x Session interaction, $F(3, 41) = .47$, $p = .71$, $\eta^2_G = .00$.

Raw Task Load Index (RTLX)

The RTLX was computed as the average of drivers' ratings on the six workload dimensions of the NASA-TLX. It could range from 0 (low workload) to 100 (high workload). Drivers answered the NASA-TLX once after Session 1 and once after Session 2. The results are depicted in Figure 40 (for group means and standard deviations refer to Table I7 in Appendix I).

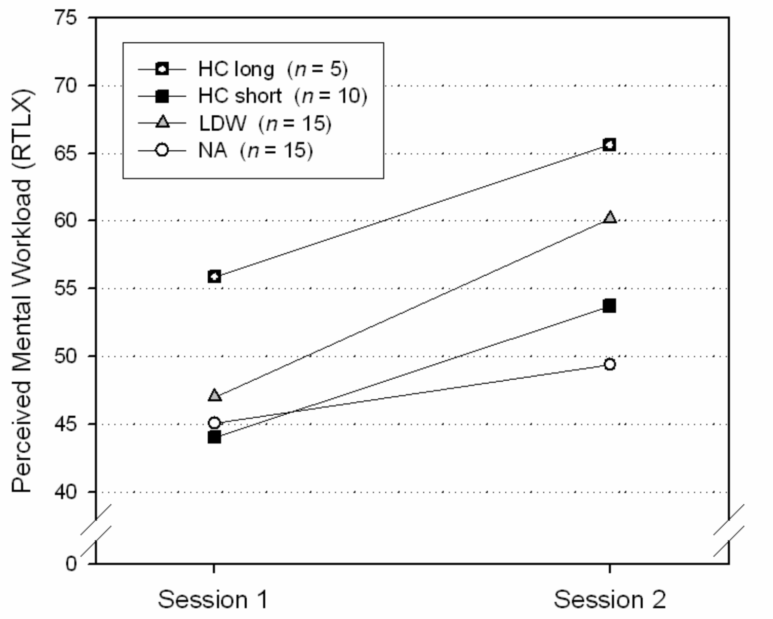


Figure 40. Drivers' ratings of overall mental workload as measured by the Raw Task Load Index (RTLX) in Session 1 and Session 2 across the three levels of lane keeping assistance, separately for the two subgroups of HC drivers that were found to differ significantly in their reliance on the HC system. The increase in drivers' overall mental workload from Session 1 to Session 2 is highly significant.

The results of the RTLX measure of drivers' overall mental workload showed a similar pattern as that obtained by the RSME. There was again a clear increase in drivers' perceived workload from Session 1 to Session 2 as response to the increase in driving task demands in Session 2 resulting from the unexpected occurrence of critical driving situations. The effect of the level of lane keeping assistance on drivers' mental workload seems to be somewhat more pronounced for the RTLX measure than for the RSME. Thus, contrary to the assumptions, NA drivers and drivers in 'HC short' subgroup generally reported a lower level of mental workload than LDW drivers and drivers in the 'HC long' subgroup. Drivers in the 'HC long' subgroup reported a considerably higher level of workload than drivers in the 'HC short' subgroup in Session 1, $t(13) = 1.57$, $p = .14$ (two-tailed), $r = .40$, and in Session 2, $t(13) = 1.69$, $p = .11$ (two-tailed), $r = .42$.

An ANOVA with session as the repeated-measures factor and level of lane keeping assistance as between-subjects factor (with four levels: HC long, HC short, NA and LDW) revealed that the main effect of the level of lane keeping assistance was (although larger) again not significant, $F(3, 41) = 2.20$, $p = .10$, $\eta^2_G = .12$. There was again no significant Level of lane keeping assistance x Session interaction, $F(3, 41) = 1.74$, $p = .18$, $\eta^2_G = .02$. As for the RSME, the effect of session became highly significant, $F(1, 41) = 27.48$, $p < .001$, $r = .32$.

Dimensions of workload

The use of the NASA-TLX as a measure of workload enabled not only to derive an overall score of drivers' workload, but also to explore further the primary contributors to this overall workload score by referring to drivers' ratings on the six subscales that reflect different workload dimensions. Figure 41 and Figure 42 present the mean workload ratings for each of the six NASA-TLX dimensions separately for the two subgroups of HC drivers as well as for LDW and NA drivers. Figure 41 depicts the workload ratings for Session 1, whereas Figure 42 depicts the workload ratings for Session 2. Group means and standard deviations can be found in Table I8 in Appendix I.

It becomes apparent in the two figures that drivers in 'HC long' subgroup show a workload signature that appears to be distinct from all other drivers, especially in Session 1. In Session 2, mainly two things become evident: First, workload ratings increase over all six dimensions (reflecting the significant increase in overall workload from Session 1 to Session 2 as measured by the RSME and RTLX), and second, variability across the four groups of drivers increases. However, also in Session 2, drivers in the 'HC long' subgroup show a somewhat distinct workload pattern. Specifically, drivers in the 'HC long' subgroup had much higher ratings of frustration, but also of temporal demands, effort, and mental demands than drivers in the 'HC short' subgroup, LDW and NA drivers. Furthermore, drivers in the 'HC long' subgroup were more dissatisfied with their own performance, as evidenced by their lower scores on the 'performance' subscale.

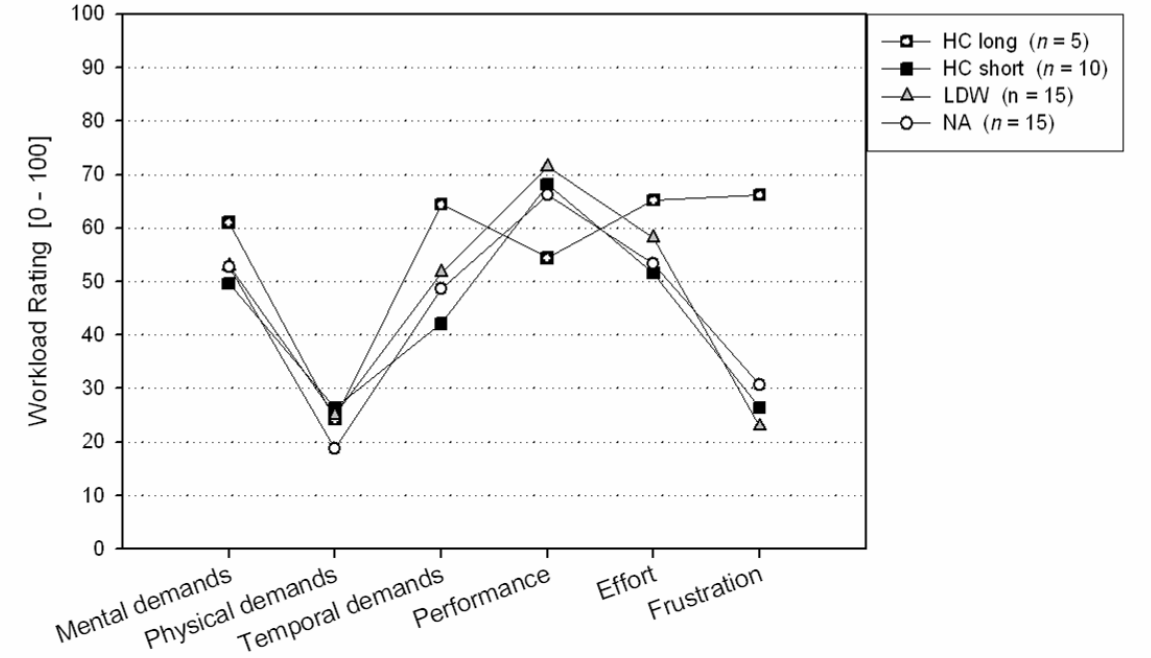


Figure 41. Drivers' workload ratings on six dimensions of mental workload as measured by the NASA Task Load Index (NASA-TLX) in Session 1 for the two subgroups of HC drivers that were found to differ significantly in their reliance on the HC system and for LDW and NA drivers. Drivers differed significantly in their ratings of frustration.

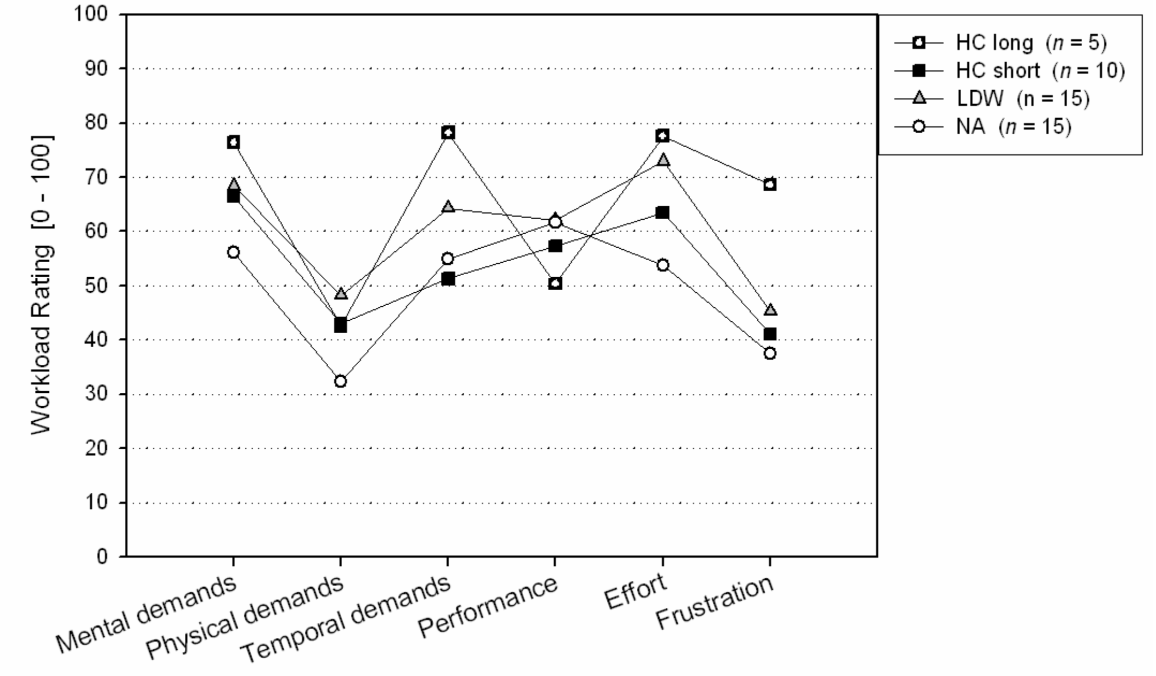


Figure 42. Drivers' workload ratings on six dimensions of mental workload as measured by the NASA Task Load Index (NASA-TLX) in Session 2 for the two subgroups of HC drivers that were found to differ significantly in their reliance on the HC system and for LDW and NA drivers. Drivers differed significantly in their ratings of frustration.

Separate analyses of variance for each workload dimension revealed that the four groups of drivers differed only significantly in their frustration ratings, $F(3, 41) = 3.72$, $p = .02$, $\eta^2_G = .17$. Post-hoc comparisons confirmed that drivers in the 'HC long' subgroup experienced a significant higher level of frustration than LDW drivers, $F(1, 41) = 9.37$, $p = .004$, $r = .43$, NA drivers, $F(1, 41) = 9.36$, $p = .004$, $r = .43$, and drivers in the 'HC short' subgroup, $F(1, 41) = 8.51$, $p = .006$, $r = .41$ (using Bonferroni-corrected α-levels).

For any other of the six workload dimensions the main effect of the level of lane keeping assistance or the Level of lane keeping assistance x Session interaction became significant. However, there was a significant effect of session for all six workload dimensions, confirming the increase in overall workload from Session 1 to Session 2. The analysis of variance tables for the six NASA-TLX subscales can be found in Appendix I (see Tables I9 – I14).

4.4.1.3 Drivers' mental effort regulation strategies

Based on theoretical frameworks it is assumed that mental effort regulation is a cognitive mechanism whereby drivers assign processing capacities to concurrent tasks based on the relative value (importance) of these tasks. Thus, it is a basic assumption that drivers *strategically* manage their effort investment based on their prioritisation of different goals. In order to get a more detailed picture about drivers' motives and goals that may have determined their mental effort investment, several questionnaire items were developed that drivers answered after both experimental session.

Drivers' relative prioritisation of the driving task and the secondary task

One of these items asked drivers to mark on a straight line how much of their overall mental effort they had been investing in the driving task vs. in the Arrows Task during Session 1 and Session 2. The results are depicted in Figure 43 (group means and standard deviations are presented in Table I15 in Appendix I).

As predicted, drivers invested considerably more effort in the driving task in Session 2 than in Session 1 as a response their experience of critical driving situations in Session 2. While drivers reported to have invested about as much of their mental effort in the driving task as in the Arrows Task in Session 1, they reported to have invested a substantially higher proportion of their mental effort in the driving task than in the Arrows Task in Session 2, demonstrating a clear shift in task priorities as a response to the increase in driving task demands in Session 2.

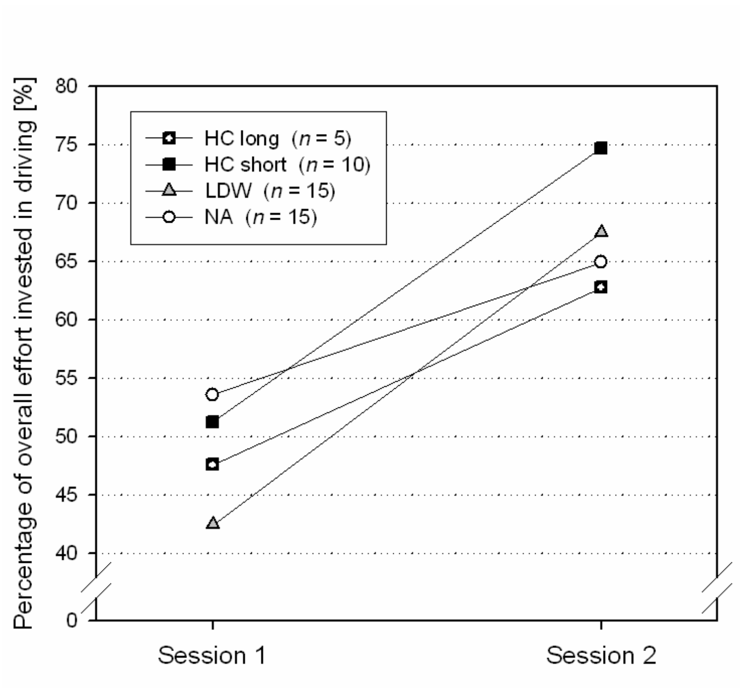


Figure 43. Mean percentage of overall effort that drivers invested in the driving task in Session 1 and Session 2 as a function of the level of lane keeping assistance, separately for the two subgroups of HC drivers. The increase from Session 1 to Session 2 is highly significant.

It was furthermore assumed that HC drivers would have invested the least mental effort in the driving task in Session 1, whereas LDW drivers should have invested a higher proportion of their mental effort in driving. NA drivers were expected to have invested the highest mental effort in the driving task in Session 1. The results partially support this prediction. NA drivers indeed reported to have invested a higher proportion of their mental effort in driving than drivers driving with a lane keeping assistance system. However, this difference was not very pronounced. On the other hand, interestingly, LDW drivers and not HC drivers reported to have invested the least mental effort in the driving task in Session 1.

HC drivers were assumed to be most affected by their experiences of functional system limits, and were therefore hypothesised to invest more mental effort in the driving task than LDW and NA drivers in Session 2. This was however only the case for the subgroup of HC drivers that had been found to rely not at all on the HC system ('HC short'), and not for the 'HC long' subgroup that had been found to rely to a large extent on the HC system. Drivers in the 'HC short' subgroup reported to have invested a considerably higher proportion of mental effort in the driving task than drivers in the 'HC long' subgroup in Session 2, $t(13) = -1.82$, $p = .09$ (two-tailed), $r = -.45$.

An ANOVA with the level of lane keeping assistance as between-subjects factor (with four levels: HC long, HC short, LDW, and NA) and session as within-subjects factor revealed a significant main effect of session, confirming that drivers invested a significant higher proportion of their mental effort in the driving task in Session 2 compared to Session 1, $F(1, 41) = 29.69$, $p < .001$, $r = .44$. Neither the main effect of the level of lane keeping assistance, $F(3, 41) = .66$, $p = .58$, $\eta^2_G = .03$, nor the interaction between level of lane keeping assistance and session were significant, $F(3, 41) = 1.29$, $p = .29$, $\eta^2_G = .03$.

Drivers' motivation to protect their driving performance

A second item asked drivers to mark on a 7-point rating scale how important it was for them that the Arrows Task did not have a negative effect on their driving performance. Lower scores on that item reflect drivers' reduced care about potential decrements in their driving performance due to the concurrent performance of the Arrows Task.

The results of drivers' ratings for that item are depicted in Figure 44 (for group means and standard deviations refer to Table I16 in Appendix I). Exploratory data analysis revealed that the two subgroups of HC drivers did not differ with respect to their ratings for that item. Therefore, they were *not* treated separately in the analysis.

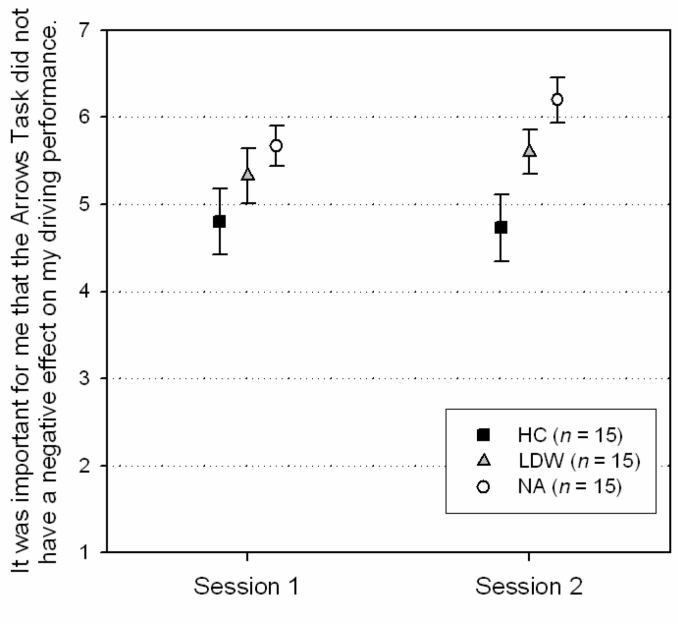


Figure 44. Drivers' motivation to protect their driving performance from deteriorations due to their concurrent performance of the Arrows Task as a function of the level of lane keeping assistance and session. The main effect of the level of lane keeping assistance is significant. Error bars depict standard errors.

As Figure 44 shows, drivers generally tried to a rather high extent to protect their driving performance from possible deteriorations due to the secondary task. However, it becomes apparent that with increasing automation of the lateral control task drivers cared less about the negative effects of the Arrows Task on their driving performance. Thus, drivers seem to feel less accountable for their driving performance the higher the level of lane keeping assistance.

An ANOVA with the level of lane keeping assistance as between-subjects factor and session as within-subjects factor revealed a significant main effect of the level of lane keeping assistance, $F(2, 42) = 4.21$, $p = .02$, $\eta^2_G = .14$. Post-hoc comparisons with Bonferroni-corrected α-levels revealed that HC drivers differed significantly from NA drivers, $F(1, 42) = 8.31$, $p = .006$, $r = .41$, but not from LDW drivers.

There was no significant main effect of session, $F(1, 42) = 3.00$, $p = .09$, $\eta^2_G = .01$, nor a significant interaction between the level of lane keeping assistance and session, $F(2, 42) = 1.51$, $p = .23$, $\eta^2_G = .01$.

Drivers' motivation to perform well on the Arrows Task

In order to find out whether the four groups of drivers differed in their motivation to perform well on the Arrows Task, drivers' ratings on a 7-point rating scale asking for their willingness to succeed on the Arrows Task were analysed. Figure 45 depicts the results. Group means and standard deviations can be found in Table I17 in Appendix I.

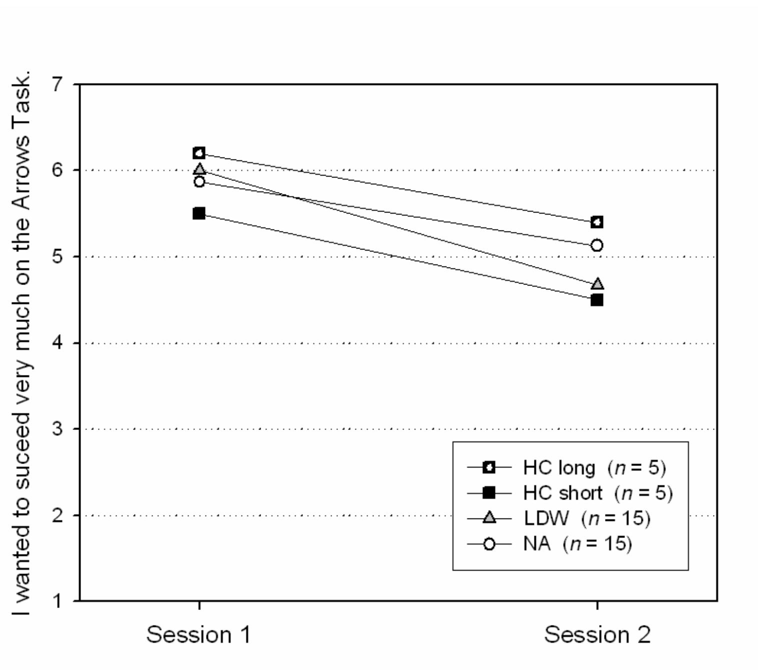


Figure 45. Drivers' motivation to perform well on the Arrows Task in Session 1 and in Session 2 as a function of the level of lane keeping assistance, separately for the two subgroups of HC drivers. Drivers' motivation significantly decreased from Session 1 to Session 2.

As can be seen in Figure 45, drivers generally had a very high motivation to perform well on the Arrows Task, strengthening the assumption that drivers did not regard the Arrows Task as a task of low priority. Interestingly, drivers' motivation to succeed well on the Arrows Task decreased from Session 1 to Session 2, reconfirming a shift in task priorities in response to the increase in driving task demands in Session 2.

Nevertheless, drivers were still highly motivated to perform well on the Arrows Task also in Session 2.

It is also interesting to note that drivers in the 'HC long' subgroup who were found to rely to a high extent on the HC system reported the highest motivation to perform well on the Arrows Task. On the contrary, drivers in the 'HC short' subgroup who were found to not rely on the HC system at all reported the lowest motivation to perform well on the Arrows Task. This effect was of medium size both in Session 1, $t(13) = 1.18$, $p = .26$ (two-tailed), $r = .31$, and in Session 2, $t(13) = 1.13$, $p = .28$ (two-tailed), $r = .29$.

An ANOVA with the level of lane keeping assistance as the between-subjects factor (HC long, HC short, LDW and NA) and session as within-subjects factor revealed a significant main effect of session, $F(1, 41) = 11.69$, $p = .001$, $\eta^2_G = .10$. Both the main effect of the level of lane keeping assistance, $F(3, 41) = .74$, $p = .54$, $\eta^2_G = .03$, as well as the Level of lane keeping assistance x Session interaction were not significant, $F(3, 41) = .33$, $p = .80$, $\eta^2_G = .01$.

Summary and discussion

The magnitude of compensatory costs was assessed as a second indirect measure of drivers' mental effort regulation strategies (besides the evaluation of subsidiary task failures). Compensatory costs refer to the physiological and affective side effects of human regulatory behaviour and constitute a form of latent performance decrement under high workload conditions. Compensatory costs were assessed by subjective measure of mental workload and frustration.

It was predicted that the magnitude of compensatory costs would increase from Session 1 to Session 2 in response to the increase in driving task demands due to drivers' unexpected encountering of critical driving situations in Session 2 that demanded their reaction in order to avoid potential collisions with other traffic. The results provided clear and consistent support for this hypothesis. There was a medium effect for the increase in drivers' perceived workload from Session 1 to Session 2 as measured by the Rating Scale Mental Effort (RSME) and the Raw Task Load Index (RTLX), $r = .32$ (for both measures).

Furthermore, other subjective measures that focused to a larger extent on drivers' mental effort regulation strategies provided clear evidence that drivers responded to the increasing driving task demands in Session 2 by a shift in task priorities. Whereas drivers rated to have assigned about equal priority to the driving task and to the Arrows Task in Session 1, their proportion of overall effort invested in the driving task markedly increased from Session 1 to Session 2, an effect that was found to be quite large ($r = .44$). The shift in task priorities from Session 1 to Session 2 became also evident in drivers' reported willingness to succeed on the Arrows Task. Drivers reported a lower motivation to perform well on the Arrows Task in Session 2 compared to Session 1 ($r = .31$).

Nevertheless, drivers reported still a high motivation to perform well on the Arrows Task in Session 2, reconfirming the assumption (which was derived from the results for drivers' Arrows Task performance) that drivers did not regard the Arrows

Task as a task of low priority. The results on drivers' mental effort regulation argue again for the notion that drivers' time-sharing strategies were somewhat less effective in Session 1, and turned to be more effective in Session 2. Remember that drivers' Arrows Task performance remained stable from Session 1 to normal driving periods in Session 2, although drivers allocated substantially less visual attention to the Arrows Task in Session 2. The increase in time-sharing "effectiveness" was however accompanied by an increase in compensatory costs (although it cannot be completely ruled out that the increase in drivers' perceived workload in Session 2 was caused mainly by the occurrence of critical driving situations because workload ratings were not collected separately for normal driving periods in Session 2 and for critical driving situations in Session 2 as it was done for the objective measures).

A second major assumption was that drivers' reliance on a lane keeping assistance system is influenced by their strategic decisions about the investment of mental effort. Specifically it was hypothesised that increasing levels of drivers' reliance on a lane keeping assistance system would (partially) result from drivers' attempt to limit their expenditure of mental effort in the driving task. Consequently, it was assumed that higher reliance on a lane keeping assistance system would be accompanied by drivers' reduced investment of driving-task related effort, which should in turn be reflected by a lower magnitude of compensatory costs as an indirect measure of drivers' effort expenditure.

Results did not support this prediction. Drivers who had been found to rely to a high extent on the HC system ('HC long' subgroup) reported the highest level of mental workload, whereas NA drivers (who were driving without lane keeping assistance) reported the lowest level of mental workload, followed by HC drivers who had been found to rely not at all on the HC system ('HC short' subgroup). Although these differences were not significant, they definitely pointed in the reverse direction.

Also, there was no evidence that HC drivers had invested substantially less mental effort in the driving task as a function of the higher level of lane keeping assistance, as it was hypothesised. Although there was a tendency that HC drivers invested proportionally less mental effort in the driving task compared to NA drivers in Session 1 and that drivers in the 'HC long' subgroup invested proportionally less mental effort in the driving task compared to drivers in the 'HC short' subgroup in Session 1, these differences were not very pronounced and not significant. There was also no clear evidence that HC drivers invested considerably more effort in the driving task than LDW and NA drivers in Session 2 in response to their experience of critical driving situations and functional limits of the HC system. Only drivers in the 'HC short' subgroup reported to have invested more mental effort in the driving task than LDW and NA drivers in Session 2, however, again these differences were not very pronounced. It has to be noted however that the ratings of the *proportion* of mental effort invested in the driving task vs. in the Arrows Task may have been insensitive to variations in the *absolute* magnitude of mental effort investment in the driving task since a decline in effort invested in driving may not automatically imply an increase in effort investment in the Arrows Task, although this constitutes the rationale of the secondary task paradigm.

A particularly intriguing finding was the higher workload ratings of the 'HC long' subgroup. The higher workload experienced by drivers in the 'HC long' subgroup indeed speak against the notion that this subgroup might have taken more time to respond to the Arrows Task because they felt less stressed and more relaxed. Furthermore, a more detailed analysis of the different workload dimensions revealed that drivers in the 'HC long' subgroup showed a workload profile that was distinct from the other HC drivers, and also from LDW and NA drivers. Specifically, drivers in the 'HC long' subgroup reported a significantly higher level of frustration, and also reported higher levels of effort, temporal demands, and mental demands. Also, drivers in the 'HC long' subgroup were more dissatisfied with their own performance than drivers in the other groups.

This pattern of results is highly inconsistent with the hypothesis that drivers' increasing reliance on a lane keeping assistance system is guided by a greater adoption of a mental effort conservation strategy whereby drivers attempt to limit the expenditure of mental effort in the driving task. The workload signature of drivers in the 'HC long' subgroup rather seems to indicate that their high level of reliance on the HC system may have been for some reason involuntary, or at least influenced by other factors beside the hypothesised motivation to use the system's support as a mean to limit mental effort investment. The behavioural pattern of the 'HC long' subgroup let suggest that their high reliance on the HC system, i.e. their long single glance durations away from the road scene, in fact may have *contributed* to their high level of frustration. Indeed, there was a strong positive correlation between drivers' reliance on the HC system (as measured by the mean duration of HC drivers' single glances to the Arrows Task display) and their frustration in Session 1, $r(13) = .59$, $p = .02$ (two-tailed). In Session 2, the correlation was weaker, $r(13) = .21$, $p = .46$ (two-tailed).

There may be other explanations for the 'HC long' subgroup's high level of perceived workload and frustration. One alternative might be that those drivers were higher motivated to perform well on the Arrows Task, which provoked them to look for longer times away from the road scene, to examine the visual display of the Arrows Task more thoroughly, and to respond to the Arrows Task not until they were completely sure about the correctness of their responses. This strategy could have resulted indeed in higher level of perceived workload, but maybe not in a higher level of frustration. Drivers in the 'HC long' subgroup indeed reported that they were more motivated than drivers in the 'HC short' subgroup to perform well on the Arrows Task; however, this effect was neither large nor significant. Furthermore, drivers in the 'HC long' subgroup showed together with drivers in the 'HC short' subgroup the worst response accuracy, which argues against this interpretation of results. If drivers in the 'HC long' subgroup indeed tried so hard to perform well on the Arrows Task, they should have shown at least a similarly high response accuracy as the control group (NA drivers).

It seems also implausible to assume that the 'HC long' subgroup experienced a higher level of workload because they had tried harder to protect their driving performance from deteriorations due to the concurrent performance of the Arrows Task. In fact, HC drivers were found to care significantly less about possible deteriorations in driving performance due to the Arrows Task than NA drivers.

A third explanation for the 'HC long' subgroup's pattern of behaviour may be that those drivers felt especially uncomfortable with the HC system's support and had problems to come to term with the HC system's steering torques, and thus felt especially hindered and distracted from performing the Arrows Task appropriately due to the HC system's interventions in lateral control.

It will be interesting to see whether the following analyses of drivers' trust in the lane keeping assistance systems, their evaluation of the systems' performance as well as of their energetic state (arousal) will provide further support for those explanations.

4.4.2 Drivers' trust in the lane keeping assistance systems

Drivers' reliance on a lane keeping assistance system was assumed to be not only influenced by motivational factors (i.e. drivers' strategic investment of mental effort), but also by drivers' attitudes towards the system. In line with Lee and See's (2004) theoretical framework drivers' trust in a lane keeping assistance system was considered as an attitude that guides their intention to rely on it. It was assumed that a higher level of reliance on a lane keeping assistance system is accompanied by a higher level of trust in the system.

Trust in automation was found to be largely influenced by the performance of an automated system; i.e. operators' trust in a system generally increased with increasing system reliability, whereas trust decreased in response to system failures or malfunctions. Furthermore, Muir and Moray (1996) found that a system's competence was a strong predictor of operators' trust in it.

Based on empirical evidence that a system's capability and performance has an important influence on operators' trust in it, it was assumed that drivers would trust the HC system to a higher extent than the LDW system based on the higher competence of the HC system to prevent them from unintentional lane departures during their concurrent performance of the Arrows Task. Thus, it was assumed that HC drivers report a higher level of trust than LDW drivers in Session 1, which in turn was hypothesised to contribute to their higher level of reliance. The results on drivers' reliance on the lane keeping assistance systems had revealed however that drivers differed considerably in their reliance on the HC system. Specifically, it was found that only one subgroup of HC drivers ('HC long' subgroup) showed a much higher level of reliance than LDW drivers, whereas the other subgroup of HC drivers ('HC short' subgroup) allocated less control to the system than LDW drivers. It was predicted therefore that – if drivers' reliance on the lane keeping assistance systems in Session 1 varied as a function of their trust in it – only drivers in the 'HC long' subgroup express a higher level of trust than LDW drivers, whereas drivers in the 'HC short' subgroup should express a lower level of trust than LDW drivers in Session 1.

Drivers' trust in the HC and LDW system was measured on a continuous scale ranging from 0 (corresponding to no trust at all) to 100 (corresponding to extremely high trust). The results for the two subgroups of HC drivers and for LDW drivers are presented in Figure 46.

As shown in Figure 46, the two subgroups of HC drivers who had been found to differ substantially in their reliance on the HC system did also differ to a considerable degree in their trust in the HC system, especially in Session 1. According to the predictions, indeed only drivers in the 'HC long' subgroup (*M* = 80.6, *SD* = 12.58) reported a higher level of trust than LDW drivers (*M* = 57.13, *SD* = 24.61) in Session 1, whereas LDW drivers and drivers in the 'HC short' subgroup (*M* = 56.9, *SD* = 24.85) did not differ in their trust ratings. Thus, although drivers in the 'HC short' subgroup were found to allocate substantially less control to the system than LDW drivers, their difference in reliance was not mirrored by a difference in their trust in the lane keeping assistance systems. A planned contrast revealed that drivers in the 'HC long' subgroup reported a significant higher level of trust than LDW drivers, *t*(27) = 1.95, *p* = .03 (one-tailed), *r* = .43, as well as than drivers in the 'HC short' subgroup in Session 1, *t*(27) = 1.86, *p* = .04 (one-tailed), *r* = .48.

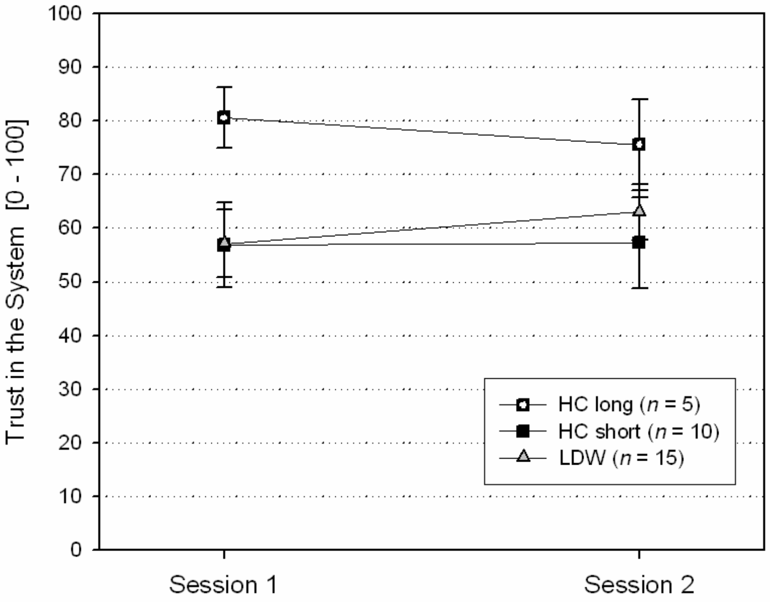


Figure 46. Drivers' trust in lane keeping assistance systems in Session 1 and in Session 2 for LDW drivers and for the two subgroups of HC drivers that were found to differ substantially in their reliance on the HC system. Error bars depict standard errors.

It was further predicted that drivers' experience of functional limits of the lane keeping assistance systems in Session 2 would have a similar effect on their trust as system failures had in previous automation studies. Thus, it was assumed that drivers' trust would decline from Session 1 to Session 2. This decline in trust was hypothesised to be more pronounced for HC drivers as a function of the HC system's greater intervention in lateral control and the more severe consequences on performance and safety when its support was inadequate. Unless drivers did not override or switch off the HC system in those situations, the HC system's actions would have resulted in collisions with other traffic. Thus, it was assumed that HC drivers would show a larger decline in trust in response to their experience of functional system limits than LDW drivers.

The data did not support this prediction. Contrary to the assumptions, there was no general decline in drivers' trust from Session 1 to Session 2, $t(27) = .11$, $p = .46$ (one-tailed), $r = .05$. Trust only declined to some extent for drivers in the 'HC long' subgroup from Session 1 to Session 2 ($M = 75.6$, $SD = 18.95$), whereas trust did not change for drivers in the 'HC short' subgroup from Session 1 to Session 2 ($M = 57.3$, $SD = 26.78$). Interestingly, LDW drivers' trust in the LDW system instead slightly increased from Session 1 to Session 2 ($M = 63.07$, $SD = 19.98$). The decline in trust from Session 1 to Session 2 was not significantly larger for drivers in the 'HC long' subgroup than for drivers in the 'HC short' subgroup, $t(27) = .51$, $p = .31$ (one-tailed), $r = .12$. Drivers were found to not differ significantly in their trust ratings in Session 2, $F(2, 27) = 1.12$, $p = .34$, $\eta^2_G = .14$.

4.4.2.1 Influences of system properties on drivers' trust in the lane keeping assistance systems

Additionally to the assessment of drivers' overall level of trust in the HC and in the LDW system, drivers were asked to evaluate specific properties of the lane keeping assistance systems that were assumed to constitute the basic information upon which trust is built (Lee & See, 2004; Muir & Moray, 1996). The lane keeping assistance systems' performance was expected to be an important determinant of drivers' trust in them. Drivers' perception of the HC and the LDW system's performance was assessed by asking them to rate the systems' competence (in preventing them from unintentional departures from the driving lane) and the predictability of the systems' actions. Furthermore, it is assumed that the way how a system acts influences operators' trust in it. Therefore, drivers were also asked to rate the comprehensibility of the systems' behaviour.

Drivers' evaluation of system competence, predictability and comprehensibility were assessed on three continuous rating scales (ranging from 0 to 100) equivalent to the trust rating scale.

In a first step it was examined whether the two subgroups of HC drivers differed substantially in their evaluation of the HC system's performance and behaviour. Independent t-tests were conducted in order to compare the two HC subgroups in their ratings of system competence, comprehensibility and predictability in Session 1 and Session 2. Only when the difference between the two subgroups was substantially large ($r \geq .3$, corresponding to a medium effect size) in at least one of the both sessions, they were treated separately in the subsequent analyses. As Table 12 shows, the two HC subgroups differed considerably in their ratings of system competence ($r = .40$) and comprehensibility ($r = .39$) in Session 1, but not in their ratings of predictability. Consequently, the two HC subgroups were treated separately in the analyses of drivers' evaluations of system competence and comprehensibility.

Table 12. Results of independent t-tests of the difference in the two HC subgroups' evaluation of the HC system's properties in Session 1 and Session 2.

	Competence			Comprehensibility			Predictability		
	$t(13)$	p^a	r	$t(13)$	p^a	r	$t(13)$	p^a	r
Session 1	1.58	.14	.40	-1.53	.15	-.39	$.65^b$	.53	.14
Session 2	.49	.63	.14	.26	.80	.07	.34	.74	.09

[a]two-tailed. [b]*df* = 12.545, *t* adjusted for unequal variances between groups.

Figure 47 presents the results of drivers' ratings of system competence, comprehensibility and predictability for Session 1 and for Session 2. The two HC subgroups are represented by separate lines for all three system properties for the reason of continuity. Group means and standard deviations can be found in Table J1 in Appendix J.

As it was already revealed in Table 12, the two subgroups of drivers differed substantially in their ratings of the HC system's competence. As it can be seen in Figure 47, drivers in the 'HC long' subgroup who had been found to have a higher trust in the HC system than drivers in the 'HC short' subgroup especially in Session 1, perceived the HC system also as being much more competent in preventing them from departures from the driving lane than drivers in the 'HC short' subgroup.

Interestingly, drivers in the 'HC long' subgroup however perceived the HC system's behaviour as less comprehensible than drivers in the 'HC short' subgroup, especially in Session 1. As Figure 47 shows, drivers generally found that the lane keeping assistance systems' actions and behaviour were to a high degree comprehensible. However, all drivers except drivers in the 'HC long' subgroup rated the systems' comprehensibility higher than their actual performance, i.e. the systems' competence in actually preventing them from departures from the driving lane. Only drivers who had been found to rely to a high extent on the HC system ('HC long' subgroup) judged the HC system's competence about equally high as the comprehensibility of its actions.

Although drivers generally perceived the systems' actions as being highly comprehensible, they rated themselves to be less able to predict the systems' actions on a moment-to-moment basis.

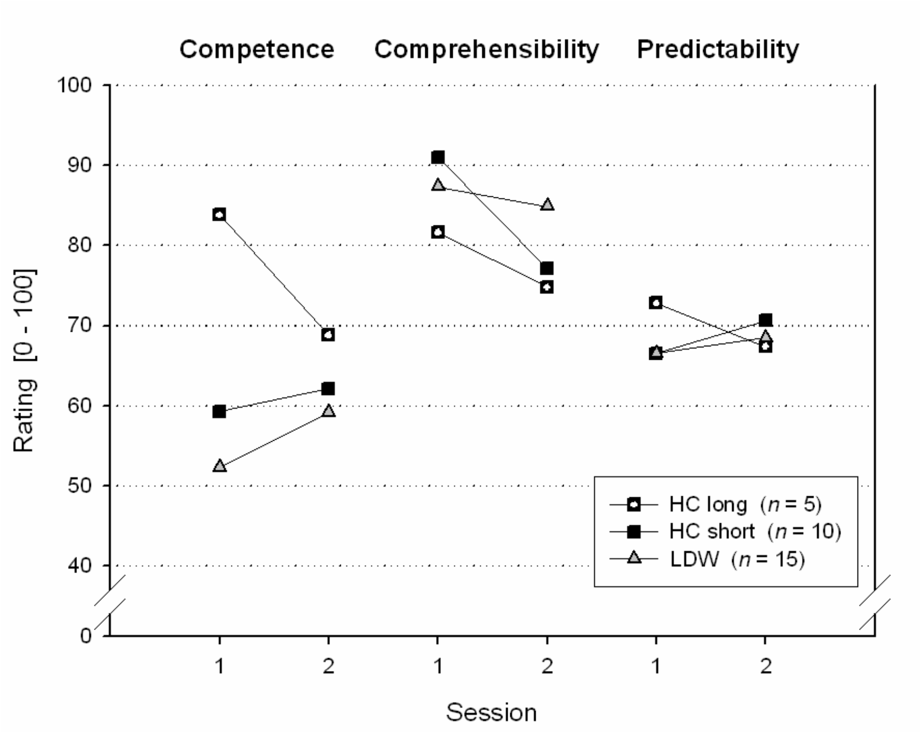


Figure 47. Drivers' evaluation of the lane keeping assistance systems' competence, comprehensibility and predictability as a function of session for LDW drivers and the two subgroups of HC drivers that differed substantially in their trust in and in their reliance on the HC system.

In the following, separate ANOVAs were conducted for drivers' ratings of system competence and comprehensibility with the level of lane keeping assistance treated as between-subjects factor (with three levels: HC long, HC short, and LDW) and session as within-subjects factor. A visual exploration of the data of drivers' evaluation of the predictability of the systems' behaviour indicated that there were no remarkable differences between HC and LDW drivers, and also no apparent effect of session. Because an ANOVA was supposed to reveal any important result, no statistical test was conducted in order to analyse drivers' ratings of predictability.

Evaluation of the lane keeping assistance systems' competence

It could have been expected that drivers perceived the HC system as being more capable than the LDW system in preventing them from departures from the driving lane because of its higher degree of intervention in lane keeping. As Figure 47 shows, this was only the case for drivers in the 'HC long' subgroup who had been found to rely to a large degree on the HC system. LDW drivers and drivers in the 'HC short' subgroup however differed not much in their perception of the HC and LDW systems' competence. Interestingly, in response to their experience of functional system limits, drivers' perception of the systems' competence declined from Session 1 to Session 2, but only for drivers who had been found to rely to a large extent on the HC system ('HC long' subgroup). On the contrary, drivers in the 'HC short' subgroup and LDW drivers did not perceive the systems' competence to be diminished due to the occurrence of functional system limits. Actually, their ratings of the systems' competence even slightly increased from Session 1 to Session 2. An ANOVA revealed that this interaction was not significant, $F(2, 27) = 1.13$, $p = .34$, $\eta^2_G = .02$. Also, there was neither a significant main effect of the level of lane keeping assistance, $F(2, 27) = 1.46$, $p = .25$, $\eta^2_G = .07$, nor a significant main effect of session, $F(1, 27) = .10$, $p = .76$, $\eta^2_G = .00$.

Evaluation of the lane keeping assistance systems' comprehensibility

The most apparent effect with regard to drivers' evaluation of system comprehensibility was that the lane keeping assistance systems' actions were perceived as less comprehensible in Session 2 in comparison to Session 1, which was probably due to the inadequate performance of the HC and the LDW system in some of the critical driving situations in Session 2. An ANOVA confirmed that the main effect of session was significant, $F(1, 27) = 7.39$, $p = .01$, $\eta^2_G = .07$. There was neither a significant main effect of the level of lane keeping assistance, $F(2, 27) = .83$, $p = .45$, $\eta^2_G = .04$, nor a significant interaction between the level of lane keeping assistance and session, $F(2, 27) = 1.98$, $p = .16$, $\eta^2_G = .04$.

Correlation analyses

Correlation analyses were performed in order to investigate further the interrelations between the three system properties and in order to find out how they were related to drivers' trust in the lane keeping assistance systems.

Table 13 presents the results of the correlation analyses for Session 1 separately for HC and LDW drivers. Table 14 depicts the respective results for Session 2.

A first glance at Table 13 and Table 14 shows that intercorrelations between the three system properties were considerably higher in Session 2 than in Session 1. Also, there are noticeable differences between HC and LDW drivers especially with regard to the correlations of drivers' evaluation of the systems' properties with their trust in the systems.

In Session 1, there was a large and significant positive correlation between drivers' perception of the HC system's competence and their trust in the HC system, $r(12) = .81$, $p < .001$. Thus, the more competent the HC system was perceived by drivers, the more did they trust it. The comprehensibility and the predictability of the HC system's behaviour were less strongly related with drivers' trust in the HC system in Session 1.

On the contrary, for LDW drivers, none of the three aspects of system behaviour and performance were significantly related to their trust in the LDW system.

In Session 2, this picture changed. Drivers' trust in the LDW system was significantly related to their perception of the predictability of the LDW system's behaviour ($r(13) = .56$, $p = .03$), whereas the LDW system's competence and comprehensibility were less strongly correlated with drivers' trust in the LDW system.

Table 13. Intercorrelations between drivers' evaluation of the lane keeping assistance systems' properties and drivers' trust in the HC and LDW system in Session 1.

Rating	1	2	3	4
HC (n = 14)[a]				
1. Competence	—	-.08	.19	.81**
2. Comprehensibility		—	.51	.27
3. Predictability			—	.40
4. Trust				—
LDW (n = 14)[b]				
1. Competence	—	-.05	-.01	.29
2. Comprehensibility		—	.25	.27
3. Predictability			—	.25
4. Trust				—

[a] one HC driver was identified as outlier in the scatterplots and therefore excluded from the analyses.
[b] one LDW driver was identified as outlier in the scatterplots and therefore excluded from the analyses.
**p < .01 (two-tailed).

For HC drivers, competence was still the system property which most strongly correlated with their trust in the HC system in Session 2. A further finding was that drivers' perception of comprehensibility and predictability of the systems' behaviour were strongly positively correlated in Session 2, both for HC drivers and for LDW drivers. In Session 1, the correlations between these two system properties were weaker, even more so for LDW drivers than for HC drivers.

For HC drivers, the HC system's competence was positively correlated with its predictability (the more predictable the HC system was experienced, the higher was its perceived competence, or vice versa), whereas no such correlation was found for LDW drivers in Session 2. For both HC and LDW drivers, their ratings of comprehensibility of the systems' behaviour were positively correlated with their ratings of the systems' competence, suggesting that drivers' understanding of the systems' behaviour positively influenced their perception of the systems' competence (although the other direction might be also possible, though less plausible).

Table 14. Intercorrelations between drivers' evaluation of the lane keeping assistance systems' properties and drivers' trust in the HC and LDW system in Session 2.

Rating	1	2	3	4
		HC (n = 15)		
1. Competence	—	.56*	.54*	.75**
2. Comprehensibility		—	.83**	.34
3. Predictability			—	.41
4. Trust				—
		LDW (n = 15)		
1. Competence	—	.52*	.01	.32
2. Comprehensibility		—	.64**	.42
3. Predictability			—	.56*
4. Trust				—

**$p < .05$ (two-tailed). *$p < .01$ (two-tailed).

Multiple regression analyses were planned in order to investigate whether drivers' trust in the HC and in the LDW system could be predicted based on their perception of the systems' competence, comprehensibility, and predictability or by a combination of these. Before these regression analyses were performed, it was investigated whether drivers' trust in the lane keeping assistance systems was influenced by individual characteristics of the drivers.

4.4.2.2 Influences of individual driver characteristics on drivers' trust in the lane keeping assistance systems

It can be hypothesised that a driver's trust in a lane keeping assistance system may not only be influenced by characteristics or properties of the system, but also by individual characteristics of the driver. Lee and See (2004) for example view operators' predisposition to trust technical systems as an important determinant of operators' trust in automation. In this study, different driver variables were assessed that might have an influence on drivers' trust in the lane keeping assistance systems. These were drivers' interest in technology, their driving style, their propensity to rely on driver assistance systems, as well as their driving experience. It can be assumed that driver characteristics especially influenced drivers' initial level of trust in the lane keeping assistance systems, whereas system properties played a more important role in the later stages of trust calibration.

Influence of driving experience on drivers' trust

In order to investigate whether drivers' driving experience was related to their trust in the lane keeping assistance systems, drivers' trust ratings were correlated with the years drivers were in possession of their driving licence. Drivers' trust in the HC system was not correlated with HC drivers' driving experience, neither in Session 1 $r(13) = .02$, $p = .94$ (two-tailed), nor in Session 2, $r(13) = .00$, $p = .99$ (two-tailed). Drivers' trust in the LDW system was however negatively correlated with LDW drivers' driving experience. The correlation was weaker in Session 1, $r(13) = -.26$, $p = .35$ (two-tailed), and larger in Session 2 where it also approached statistical significance, $r(13) = -.52$, $p = .05$ (two-tailed). Thus, the longer LDW drivers were in possession of their driving licence, the less did they trust the LDW system.

Influence of drivers' propensity to rely on driver assistance systems on their trust

Drivers' propensity to rely on driver assistance systems was assessed prior to the first experimental session on a 4-point rating scale with descriptors that ranged from an unbiased attitude towards these systems, expressed as a readiness to rely on them, over to a highly sceptical attitude towards these systems, coupled with a dislike to handle control over to them (see page 4 in Appendix B). Generally it was found that drivers' ratings on that scale did not vary very much. However, as could have been assumed, a more sceptical attitude of drivers towards driver assistance systems (corresponding to a low propensity to rely on them) was related to a lower level of trust in the LDW and in the HC system. The correlations were $r(13) = -.15$, $p = .60$ (two-tailed) and $r(13) = -.32$, $p = .27$ (two-tailed) for HC drivers in Session 1 and in Session 2, respectively, and $r(13) = -.43$, $p = .11$ (two-tailed) and $r(13) = -.21$, $p = .45$ (two-tailed) for LDW drivers in Session 1 and Session 2, respectively.

Influence of drivers' interest in technology on their trust

Drivers' interest in technology was assessed prior to the first experimental session on seven 7-point rating scales. Factor analyses (principal component analysis with Varimax rotation) were performed in order to find out whether these seven items measured the same underlying construct. The aim was to finally derive a one-factor solution with acceptable internal consistency. An initial factor analysis yielded a three-factor solution (with eigenvalues greater than 1). However, it was found that only one item ("I do not care so much about the appearance of my car. The main thing is that it takes me where I want to go.") loaded high (> .5) on the third factor. This item seemed to be measuring something different and was therefore excluded from the subsequent analyses. A second factor analysis with the aforementioned item left out yielded a two-factor solution. A reliability analysis of the three items that loaded high on factor one yielded an internal consistency of Cronbach's alpha = .67, while exclusion of any item would have resulted in a reduction of Cronbach's alpha. A reliability analysis of the three items that loaded high on the second factor yielded a Cronbach's alpha of .74, while it was indicated that exclusion of the item "I like cars because of their technical equipment." would leave Cronbach's alpha unchanged. Therefore, this item was also excluded from subsequent analyses and a new factor analysis was performed with the remaining 5 of the 7 original items. This factor analysis yielded a one-factor solution which explained a total of 52.7% of the variance. A subsequent reliability analysis for the final one-factor solution obtained a Cronbach's alpha of .77. Table 15 depicts the factor loadings of the five final items constituting the "Interest in technology" factor.

Table 15. Factor pattern loadings and communality of "interest in technology" items (*N* = 45).

Item	Factor 1 (Interest in technology)	Communality
I'm keen on testing new technical devices.	.73	.54
I like to have new technical stuff even if I don't really need it.	.75	.56
I'm more interested in how technical things work, not if they're functioning perfectly.	.77	.59
I'm not attracted by the newest technical things on the market.	-.69	.47
I do not care about how technical devices work; the main thing is that they work.	-.69	.48

In order to reveal whether drivers' trust in the lane keeping assistance systems was related to their interest in technology, drivers' individual factor scores on the "Interest in technology" factor were correlated with their trust ratings. The correlation analyses revealed that drivers' trust in the lane keeping assistance systems was positively correlated with their interest in technology. The correlation was larger for LDW drivers in Session 1, $r(13) = .47$, $p = .08$ (two-tailed), than for HC drivers in Session 1, $r(13) = .21$, $p = .44$ (two-tailed). In Session 2, the correlations between drivers' interest in technology and their trust in the LDW and in the HC system was rather small, and pointed in the reverse direction, $r(13) = .23$, $p = .40$ (two-tailed) and $r(13) = -.22$, $p = .44$ (two-tailed) for LDW and HC drivers, respectively.

Influence of driving style on drivers' trust

Driving style was assessed with ten 7-point semantic differentials with endpoints describing different driving styles. Unfortunately, the ratings of two drivers in the HC group were not available. A factor analysis was performed (principal component analysis with Varimax rotation) in order to investigate whether those ten items reflected some joint underlying dimensions. An initial factor analysis yielded a three-factor solution. It turned out that the variance of especially one item could not be very well explained by the three factors ($R^2 = .57$). This item (asking drivers to evaluate themselves as a "good" vs. a "bad" driver) was also very general in nature and assumed to be potentially subject to biases (social desirability). It was therefore excluded from the subsequent analysis. A second factor analysis with the remaining nine items again yielded a three-factor solution (with eigenvalues greater than 1) which together accounted for 73.18% of the total variance. Those three factors were labelled as "offensive vs. defensive driving", "unconfident vs. confident driving", and "observant vs. unobservant driving". The "offensive vs. defensive driving" factor contrasted fast, risk accepting, determined and offensive driving with slow, cautious, hesitant, and defensive driving. The "unconfident vs. confident driving" factor contrasted anxious, emotional, and nervous driving with courageous, calm, and confident driving. Finally, the "observant vs. unobservant driving" factor contrasted fore-sighted, observant driving with impulsive, unobservant driving. Factor loadings of the nine items are shown in Table 16.

Table 16. Factor pattern loadings and communality of driving style items for the rotated solution (*N* = 43).

Driving style	Factor 1 (offensive vs. defensive)	Factor 2 (unconfident vs. confident)	Factor 3 (observant vs. unobservant)	Communality
slow … fast	**.83**	-.02	.02	.69
cautious … risk accepting	**.62**	.06	.50	.63
hesitant … determined	**.77**	.21	-.34	.76
defensive … offensive	**.71**	.12	.34	.64
anxious … courageous	.31	**.82**	.18	.80
calm … emotional	.30	**-.69**	.40	.73
nervous … confident	.38	**.56**	-.51	.73
observant … unobservant	.01	-.09	**.88**	.78
fore-sighted … impulsive	.14	-.07	**.90**	.84

Note. Highest factor loadings for each item are in bold.

In order to find out whether drivers' trust in the lane keeping assistance systems was related to their driving style, drivers' individual scores on three driving style factors were correlated with their trust ratings. The results revealed some interesting findings (see Table 17). Drivers' trust in the HC system was strongly negatively correlated with the "unconfident vs. confident driving" factor especially in Session 1, $r(11) = -.55$, $p = .05$ (two-tailed), meaning that drivers trusted the HC system to a higher extent when their driving style could be characterised as anxious, emotional, and nervous. On the contrary, drivers' trust in the LDW system was found to be less negatively correlated with the "unconfident vs. confident" driving factor. However, for LDW drivers there was a strong positive correlation between their trust in the LDW system and their loadings on the "observant vs. unobservant" driving factor, which was significant both for Session 1, $r(13) = .68$, $p = .005$ (two-tailed), and for Session 2, $r(13) = .60$, $p = .02$ (two-tailed). Thus, the more LDW drivers' driving style could be characterised as unobservant and impulsive, the more did they trust the LDW system.

Table 17. Correlations between driving style factors and drivers' trust in the HC system and in the LDW system in Session 1 and Session 2.

	HC[a]		LDW[b]	
Driving style factor	Session 1	Session 2	Session 1	Session 2
offensive vs. defensive	-.12	.06	.07	.12
unconfident vs. confident	-.55*	-.37	-.21	-.23
observant vs. unobservant	.16	-.22	.68**	.60*

[a]$n = 13$. [b]$n = 15$.
*$p < .05$ (two-tailed). **$p < .01$ (two-tailed).

4.4.2.3 Determinants of drivers' trust in the lane keeping assistance systems

Multiple regression analyses were performed in order to investigate the relative predictive value of system properties (competence, comprehensibility, and predictability) and individual driver characteristics on drivers' trust in the HC and in the LDW system. Hierarchical multiple regression was used to evaluate which system properties had the largest influence on drivers' trust in the HC and LDW system in Session 1 and 2. Afterwards it was determined whether individual driver characteristics were able to significantly increase the proportion of variance explained by the previous models. Regression analyses were performed separately for HC and for LDW drivers, and separately for the two experimental sessions.

Drivers' trust in the HC system

Drivers' ratings of the HC system's competence and comprehensibility turned out to be the best predictors for drivers' trust in the HC system in Session 1, $F(2, 11) = 17.87$, $p < .001$[14]. Together, competence and comprehensibility accounted for 76.5% of the variance in drivers' trust, while competence had a greater influence on drivers' trust, $\beta = .84$, $t(11) = 5.69$, $p < .001$, than comprehensibility, $\beta = .33$, $t(11) = 2.27$, $p = .04$. It was tested whether individual driver variables that had been found to correlate with drivers' trust in the HC system (propensity to rely on driver assistance systems,

---

[14] One HC driver was identified as outlier in the residual analysis and therefore excluded from the analysis.

interest in technology, and driving style) would further increase the proportion of variance explained. It was found that only the "unconfident vs. confident" driving style factor further significantly contributed to the model, $\Delta F(1, 8) = 5.40$, $p = .05$[15]. The final model with system competence, comprehensibility, and "unconfident vs. confident" driving style as predictors explained 87.5% of the variance in drivers' trust in the HC system in Session 1, $F(3, 8) = 18.66$, $p = .001$, while competence still had the greatest influence on drivers' trust in the HC system, $\beta = .65$, $t(8) = 4.56$, $p = .002$, followed by comprehensibility, $\beta = .45$, $t(8) = 3.47$, $p = .008$, and "unconfident vs. confident" driving style, $\beta = -.34$, $t(8) = -2.32$, $p = .049$.

In Session 2, drivers' perception of the HC system's competence was also a better predictor for drivers' trust in the HC system than comprehensibility and predictability. However, only 54.5% of the variance in drivers' trust could be explained by their ratings of the HC system's competence, while adding either comprehensibility or predictability as predictors did not significantly increase the variance explained. There was also high multicollinearity between the three system properties, suggesting that they were not regarded as independent from each other by drivers. It was tested whether drivers' trust in the HC system in Session 1 would serve as a good predictor for drivers' trust in the HC system in Session 2. This was indeed the case. It turned out that drivers' trust in Session 1 served as the best predictor for drivers' trust in the HC system in Session 2, $F(1, 12) = 43.46$, $p < .001$, while inclusion of either any of the three system properties or individual driver variables did not significantly contribute to the prediction power of the model. Drivers' trust in the HC system in Session 1 explained 78.4% of the variance in drivers' trust in Session 2, $\beta = .89$, $t(12) = 6.59$.

Drivers' trust in the LDW system

The correlation analyses had revealed that drivers' trust in the LDW system in Session 1 was only weakly correlated with drivers' perception of the LDW system's competence, comprehensibility and predictability. The multiple regression analyses confirmed that finding that none of the three system properties could explain a significant proportion of the variance in drivers' trust in the LDW system. Out of the three system properties, comprehensibility was the variable which accounted for the highest proportion of the variance in drivers' trust in the LDW system, $R^2 = .28$, $F(1, 12) = 4.59$, $p = 0.53$[16].

Subsequent regression analyses with individual driver variables as predictors (driving experience, driving style, propensity to rely on driver assistance systems, and interest in technology) revealed that the "observant vs. unobservant" driving style factor served as the best predictor for drivers' trust in the LDW system in Session 1, $F(1, 12) = 10.11$, $p = .008$, $\beta = .68$, $t(12) = 3.18$, although it could explain only 45.7% of the variance in drivers' trust. None of the other individual driver variables did significantly increase the prediction power of the model.

---

[15] Driving style ratings were unavailable for two HC drivers.

[16] One LDW driver was identified as outlier in the residual analysis and therefore excluded from the analysis.

Although none of the three system properties was able to explain a significant proportion of the variance in drivers' trust in the LDW system in Session 1, it was tested whether this was also the case for Session 2. It turned out that out of the three system properties, only drivers' perception of the LDW system's predictability was suitable to predict drivers' trust in the LDW system in Session 2, $F(1, 12) = 8.02$, p = .015, $R^2 = .40$. As for HC drivers, it was tested whether LDW drivers' trust in the system in Session 2 could be predicted by their trust in Session 1. Again, this was the case. When drivers' trust ratings of Session 1 were entered as first predictor in the model, neither predictability nor any other variable representing individual driver characteristics did significantly increase the prediction power of the model. Drivers' trust in the LDW system in Session 1 explained 75.2% of the variance in drivers' trust in the system in Session 2, $F(1, 12) = 36.33$, $p < .001$, with $\beta = .87$, $t(12) = 6.03$.

4.4.2.4 Influence of drivers' trust in the lane keeping assistance systems on their reliance

It was assumed that drivers' reliance on the lane keeping assistance systems would be influenced by their trust in the systems. The results on drivers' trust in the HC and in the LDW system indicated that drivers' differences in reliance may indeed be explained partially by their level of trust in the systems. The two subgroups of HC drivers who had been found to differ significantly in their reliance on the HC system were also found to differ considerably in their trust in the HC system. More specifically, drivers in the 'HC long' subgroup who had been found to rely to a large extent on the HC system reported a significantly higher level of trust in the HC system than drivers in the 'HC short' subgroup (who were found to not rely on the HC system at all) in Session 1.

In order to test the hypothesis of the linear positive relationship between drivers' trust in the lane keeping assistance systems and their reliance more precisely, drivers' trust ratings were correlated with the mean duration of their single display glances during Arrows Task performance as a measure of their reliance on the lane keeping assistance systems. The correlation analyses were performed separately for HC and LDW drivers and for both experimental sessions.

The correlation analyses revealed that drivers' trust in the HC system was significantly positively correlated with their trust in the HC system, $r(13) = .47$, $p = .04$ (one-tailed), for both sessions. On the contrary, drivers' reliance on the LDW system was found to be not related to their trust in the LDW system in Session 1, $r(9) = -.002$, $p = .99$ (two-tailed); whereas a positive but comparably weak correlation was found for Session 2, $r(9) = .34$, $p = .15$ (one-tailed)[17]. Thus, drivers' trust had an important influence on drivers' reliance on the HC system, but not so much on their reliance on the LDW system.

---

[17] The eye glance behaviour of three LDW drivers could not be analysed due to insufficient tracking quality. One further LDW drivers was identified as outlier in the scatter plots and therefore excluded from the analysis.

4.4.2.5 Summary and discussion

It was assumed that drivers' reliance on the lane keeping assistance systems would not only be influenced by motivational processes, i.e. drivers' strategic investment of mental effort, but also by their attitudes towards the lane keeping assistance systems. The concept of trust in automation plays a dominant role in theoretical frameworks on operators' reliance on automation. In general, operators are assumed to rely on automation they trust, and to reject automation they distrust. In line with Lee and See's (2004) theoretical framework it was assumed that drivers' trust in a lane keeping assistance system guides their intention to rely on it.

The results indicate that drivers' trust in the system was an important influencing factor on drivers' reliance on the HC system, but not so much on the LDW system. Drivers who had been found to rely to a high extent on the HC system ('HC long' subgroup) reported a significant higher level of trust in the HC system than drivers who had been found to not rely on the HC system at all ('HC short' subgroup) in Session 1 (r = .48). Furthermore, there was a strong linear positive correlation between HC drivers' trust in the HC system and their reliance on it as measured by the mean duration of their single glances away from the road scene, both for Session 1 and for Session 2 (r = .47). Subsequent analysis of drivers' evaluation of different system properties that were assumed to constitute the basis on which operators' trust in automation is built confirmed that drivers' perception of the HC system's competence in preventing them from departures from the driving lane was the strongest predictor of their trust in it. Drivers in the 'HC long' subgroup who reported a considerably higher level of trust in the HC system than drivers in the 'HC short' subgroup in Session 1 also perceived the HC system as more competent in preventing them from departures from the driving lane than drivers in the 'HC short' subgroup (r = .40). The results support the assumption that a system's performance has an important influence on operators' trust in it. It remains however unclear why drivers in the HC group differed so much in their perception of the HC system's competence, and subsequently in their level of trust in it.

The results furthermore suggest that the level of automation had an impact on the time dynamics in the relation between drivers' trust and reliance, as well as on the determinants of drivers' trust. Whereas HC drivers' trust had a strong influence on their reliance on the HC system in both experimental sessions, LDW drivers' reliance was unaffected by their level of trust in the LDW system in Session 1 (r = -.002). In Session 2 however, there was a linear positive correlation between drivers' trust in the LDW system and their reliance on it, though this correlation was found to be less pronounced than for HC drivers (r = .34). These findings seem to indicate on the one hand that trust played a comparably less important role in influencing drivers' reliance on the LDW system, and on the other hand, that trust needs sufficient experiences of a lane keeping assistance system's behaviour in order to grow. Whereas the HC system was permanently "active" and could be experienced continuously by drivers through its interventions in lateral control, the LDW system provided drivers with a warning only in the comparably rare cases when they were about to cross the lane boundaries. Thus, drivers were able to experience the LDW system's actions less frequently than those of the HC system which in turn may have hindered them to develop a meaningful level of trust in the LDW system that subsequently guided their reliance on it. This assumption is also supported by the finding that drivers' trust in the LDW system actually slightly increased from Session 1 to Session 2 despite that drivers encountered several critical driving situations in Session 2 that simulated functional limits of the LDW system. Furthermore, multiple regression analyses revealed that none of the three system properties (competence, comprehensibility, predictability) could explain a sufficient proportion of the variance of drivers' trust in the LDW system in Session 1. Thus, drivers' trust in the LDW system was largely unaffected by the LDW system's performance in Session 1, and was in contrast more strongly influenced by individual driver characteristics.

The correlation and regression analyses further suggest that different mechanism may play a role in the development of drivers' trust in a lane keeping assistance system dependent on the level of assistance. Whereas drivers' perception of the systems' competence and comprehensibility were strong predictors of drivers' trust in the HC system in Session 1, these system properties had no significant influence on drivers' trust in the LDW system. On the contrary, drivers' perception of the LDW system's predictability served as a predictor of their trust in the LDW system at least in Session 2. In generalisation of these results it can be hypothesised that a system's competence and the comprehensibility of its behaviour play a more important role than its predictability in the development of drivers' trust in driver assistance systems that intervene to a higher degree in driving, whereas drivers' ability to predict a system's behaviour plays a more important role than a system's competence and comprehensibility in the development of drivers' trust in warning systems. More research is needed to provide further evidence for this assumption.

Interestingly, it turned out that individual driver characteristics had a significant influence on drivers' trust in the lane keeping assistance systems. It was found that among drivers' propensity to rely on driver assistance systems, their driving experience (measured by the years that drivers were in possession of their driving licence), their interest in technology and their driving style, only driving style explained a significant proportion of the variance in drivers' trust in Session 1. Furthermore, drivers' trust in the HC and in the LDW system was influenced by different facets of driving style. If drivers regarded themselves as a confident vs. unconfident driver had a significant influence on their trust in the HC system in Session 1. Thus, the more drivers described themselves as an anxious, emotional, and nervous driver (vs. as a courageous, calm, and confident driver) the more did they trust the HC system. On the other hand, whether drivers regarded themselves as an observant vs. unobservant driver explained a significant proportion of the variance in drivers' trust in the LDW system in Session 1. Thus, the more drivers described themselves as an unobservant and impulsive driver (vs. as an observant and fore-sighted driver) the more did they trust the LDW system. In addition, an unobservant vs. observant driving style turned out to be the best predictor of drivers' trust in the LDW system in Session 1. For HC drivers, the HC system's competence and its comprehensibility had a larger influence on their trust in the HC system in Session 1 than a confident vs. unconfident driving style. The finding that drivers' trust in the HC and in the LDW system was influenced by different facets of driving style may also be interpreted in the way that different mechanisms play a role in the development of trust in driver assistance systems depending on the degree to which they automate the driving task. Thus, it seems that drivers' trust in higher-level automated systems is more strongly influenced by affective components of driving style (e.g., anxious vs. courageous), whereas drivers' trust in a warning system is more strongly influenced by cognitive components of driving style (anticipating vs. reactive).

The correlation analyses revealed that intercorrelations between drivers' ratings of the three system properties were larger in Session 2 than in Session 1, especially for HC drivers, suggesting that drivers did not regard the properties as independent from each other and that drivers' evaluation of one system property influenced their perception of the other properties. This was also evidenced by a high multicollinearity between the three system properties when they were entered at one time as predictors in the regression model. There was a particular high positive correlation between drivers' perception of the systems' comprehensibility and predictability in Session 2, indicating that if a system's actions were perceived as comprehensible, they were also perceived as more predictable by drivers, or vice versa. Thus, it seems that, contrary to the classification of Lee and See (2004), a system's predictability conveys rather information about *process* (how the automated system operates) than about *performance* (what the automated system does). Furthermore, drivers' ratings of the HC system's competence were strongly positively correlated with their perception of the HC system's comprehensibility and predictability in Session 2, suggesting that the HC system was perceived as more competent the more its behaviour could be comprehended and predicted by drivers (though the other direction might also be possible, although less plausible). For the LDW system, competence was positively correlated with comprehensibility, but not with predictability. Thus, the ability of drivers to predict the LDW system's behaviour was not related to their perception of the LDW system's competence to prevent them from departures from the driving lane. In summary it can be stated that drivers draw inferences among different aspects of system behaviour and performance particularly after some time of interaction with the systems, and that drivers regard a system as more competent when they understand its behaviour. For higher-level automation, the ability to predict what the system does next also positively influences its perceived competence. (It has to be noted however that correlations do not allow for causal interpretations of relationships.)

Furthermore it turned out that drivers' trust in the lane keeping assistance systems was relatively stable over the course of the study. Although it was found that drivers' reliance on the lane keeping assistance systems declined in Session 2 in response to their experience of critical driving situations that partly represented functional system limits, their was no general decline in drivers' trust from Session 1 to Session 2 contrary to the predictions. There was a slight decline in trust for those drivers who had expressed a high level of trust in the HC system before ('HC long' subgroup), but no change in trust for drivers in the 'HC short' subgroup, and even a slight increase in trust for drivers in the LDW group. Although trust declined somewhat for drivers in the 'HC long' subgroup from Session 1 to Session 2, their trust did not fall below the level of drivers in the 'HC short' subgroup in Session 2. This finding replicates the results on drivers' reliance on the lane keeping assistance systems, where it was also found that the decline in the mean duration of drivers' single glances away from the road scene was more pronounced for drivers in the 'HC long' subgroup than for drivers in the ‘HC short’ subgroup. However, glance durations of the ‘HC long’ subgroup were still considerably longer than those of the ‘HC short’ subgroup in Session 2.

The finding that drivers’ trust did not change markedly due to drivers’ experience of functional system limits is even more interesting based on the fact that drivers perceived the actions of the lane keeping assistance systems to be significantly less comprehensible in Session 2 compared to Session 1. Also, there was a marked decline in drivers’ perception of the HC system’s competence from Session 1 to Session 2, but only for drivers in the ‘HC long’ subgroup. Thus, although the experience of functional system limits seemed to affect drivers’ perception of system properties, it did not markedly reduce their trust in the lane keeping assistance systems.

The “stability” of trust is also evidenced by the finding that drivers’ trust in Session 1 was the single best predictor of their trust in the lane keeping assistance systems in Session 2.

In conclusion, the results on drivers’ trust and underlying variables suggest that trust was an important influencing factor on drivers’ reliance on the HC system, but not that much on the LDW system. The performance of the HC system played a dominant role in the development of drivers’ trust in the HC system. Drivers’ perception of the HC system’s competence (to prevent them from departures from the driving lane) was the strongest predictor of their trust in it. There is still some lack of clarity concerning the two HC subgroups’ difference in frustration. Both a high level of trust in as well as a high level of reliance on a lane keeping assistance system was conceptually assumed to be linked with lower compensatory costs, i.e. with a low level of perceived mental workload and frustration. However, the present data indicate that the opposite seems to be the case. In the following, correlation analyses were performed in order to reveal whether drivers’ trust in the HC system was linked to drivers’ frustration, and whether frustration was linked to drivers’ perception of the HC system’s performance.

4.4.2.6 Relationship between drivers’ trust in the HC system and their frustration

In order to find out whether drivers’ trust in the HC system was related to their level of frustration, HC drivers’ trust ratings were correlated with their ratings of frustration. Indeed, there was a considerable positive correlation between drivers’ trust in the HC system and their perceived level of frustration, both for Session 1, $r(13) = .46$, $p = .08$ (two-tailed), and for Session 2, $r(13) = .44$, $p = .10$ (two-tailed). Thus, the more drivers trusted the HC system, the higher was their frustration. (Despite that no causal relation is indicated by a correlation, an interpretation of the relationship in the opposite direction seems to make less sense.)

Further correlation analyses were performed in order to investigate the nature of this relationship in more detail. More specifically, it was figured out whether HC drivers’ frustration was also related to the HC system’s performance which was found to be a major predictor of drivers’ trust in it. For this purpose, drivers’ ratings of the

HC system's competence (in preventing them from unintentional departures from the driving lane) were correlated with their ratings of frustration. The correlations turned out to be positive and highly significant for Session 1, *r*(13) = .65, *p* = .008 (two-tailed), and to be considerably strong also for Session 2, *r*(13) = .47, *p* = .07 (two-tailed). Thus, the more competent the HC system was perceived by drivers in preventing them from departures from the driving lane, the more frustrated did they feel. The latter finding indeed argues for the interpretation that the higher level of frustration of drivers in the 'HC long' subgroup may at least partially result from their dissatisfaction with their own lane keeping performance. Thus, the continuous interventions of the HC system in drivers' lateral control may have been experienced by drivers in the 'HC long' subgroup as continuous reminders that their own lane keeping performance was not good enough. This interpretation of results is further strengthened by the fact that HC drivers' ratings on the item "The system enhanced my lane keeping performance" were also positively correlated with their level of frustration in Session 1, *r*(13) = .45, *p* = .09 (two-tailed), and in Session 2, *r*(13) = .34, *p* = .21 (two-tailed). Also, there was a high positive correlation between drivers' ratings on the item "The system enabled me to pay more attention to the Arrows Task than I would have been able to do without the system" and drivers' ratings of frustration in Session 1, *r*(13) = .64, *p* = .009 (two-tailed), and in Session 2, *r*(13) = .49, *p* = .06 (two-tailed). This latter correlation clearly contradicts the hypothesis that high reliance on a lane keeping assistance system follows from drivers' adoption of an effort conservation strategy by which they partly handle over control for driving to the system in order to have more resources left for performance of the Arrows Task. Rather, it appears that drivers' impression of the HC system helping them to concentrate to a higher extent on the Arrows Task than they would have been able to do without its support was not experienced as a benefit by drivers, but as a source of frustration.

It was tested whether the positive correlation between drivers' trust in the HC system and their frustration was eliminated when the influence of the HC system's perceived competence was partialled out. The partial correlation between drivers' trust in the HC system and their frustration was indeed weaker, although not eliminated, when the effect of the HC system's competence was removed, *pr*(12) = .22, *p* = .46 (two-tailed) for Session 1, and *pr*(12) = .15, *p* = .62 (two-tailed) for Session 2. Thus, 99.8% and 99.9% of the common variance in drivers' trust in the HC system and their frustration could be explained by the HC system's competence in Session 1 and in Session 2, respectively. This argues for the notion that the HC system's support in preventing drivers from safety-critical deteriorations in lane keeping performance (departures from the driving lane) was on the one hand a major basis for their trust in the HC system, but on the other hand a reason for their frustration and supposedly the feeling that their own lane keeping performance was inadequate.

4.4.3 Energetic arousal

Energetic processes were considered as the third influencing factor on drivers' reliance on the lane keeping assistance systems. Based on empirical evidence that operators' level of activation decreased in response to the increasing reliability of an automated system it was hypothesised that increasing automation of the lateral control task may lead to a lower level of energetic arousal in drivers and to problems in sustained attention. It was further assumed that low activation impairs drivers' ability for vigilant system monitoring. Reliance on automation was found to be negatively correlated with operators' effort invested in system monitoring and verification of its actions. Thus, it was assumed that a low level of arousal fosters drivers' (uncritical) reliance on the lane keeping assistance systems.

It was hypothesised that HC drivers experience the lowest level of arousal during the practice session and Session 1, followed by LDW drivers and drivers in the control group. It was further assumed that HC drivers would react to the unexpected encountering of critical driving situations in Session 2 by a mobilisation of reserve attentional resources. Thus, HC drivers were predicted to report an increase in arousal and activation from Session 1 to Session 2. On the contrary, drivers in the control group were hypothesised to respond to the increase in driving task demands from Session 1 to Session 2 by a decline in arousal due to an exhaustion of attentional resources over time. LDW drivers were predicted to show no relevant change in arousal.

Drivers' level of activation and arousal was measured by the Stanford Sleepiness Scale (SSS). Drivers filled-in the SSS at three times during the study, one time after the practice session (where they drove without concurrently performing the Arrows Task), one time after Session 1, and a third time after Session 2. Scores could range from 1 to 7, with higher scores reflecting a lower level of arousal.

Explorative data analysis revealed that the two subgroups of HC drivers did not differ in their ratings on the SSS in the practice session and in Session 2. However, the 'HC long' subgroup expressed a lower level of activation than the 'HC short' subgroup in Session 1. The effect of this difference could be regarded as sufficiently large, $r = .31$, $t(13) = 1.16$, $p = .27$, so that the two subgroups were treated separately in the following analyses. Drivers' ratings on the SSS are depicted in Figure 48. Group means and standard deviations are presented in Table K1 in Appendix K.

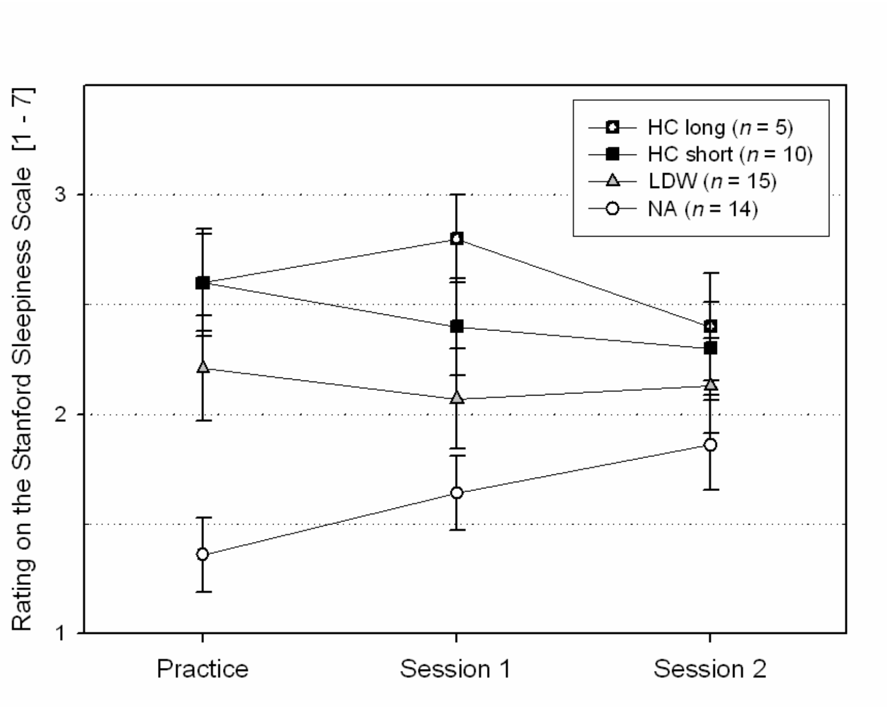


Figure 48. Drivers' arousal as measured by their ratings on the Stanford Sleepiness Scale (SSS) for the two subgroups of HC drivers and for LDW and NA drivers in the practice session (10-min drive without secondary task), in Session 1 (15-min drive with secondary task) and in Session 2 (25-min drive with secondary task and occurrence of critical driving situations). HC drivers reported a significant lower level of arousal than NA drivers. Error bars depict standard errors.

Planned contrast revealed that the decline in NA drivers' activation from the practice session over Session 1 to Session 2 was significant, $t(13) = 1.99$, $p = .03$ (one-tailed), $r = .30$. The increase in activation for drivers in the 'HC short' subgroup from the practice session to Session 1 over to Session 2 was not statistically significant, $t(9) = 1.15$, $p = .14$ (one-tailed), $r = .18$. There was no linear trend in LDW drivers' activation, $t(13) = 0.00$[18]. Planned pairwise contrasts were conducted in order to reveal whether the slopes in drivers' change in activation over the study differed in the predicted direction for LDW drivers, drivers in the 'HC short' subgroup, and NA drivers. Results indicated that the linear increase in activation from the practice session over Session 1 to Session 2 was not more pronounced for drivers in the 'HC short' subgroup than for LDW drivers, $t(39) = .79$, $p = .22$ (one-tailed), $r = .13$. The decline in activation was stronger for NA drivers than for LDW drivers but did only approach statistical significance, $t(39) = 1.45$, $p = .08$ (one-tailed), $r = .23$. Drivers in the 'HC short' subgroup showed a significantly stronger increase in arousal than NA drivers, $t(39) = 2.12$, $p = .02$ (one-tailed), $r = .32$. As shown in Figure 48, drivers in the 'HC long' subgroup differed in their activation pattern over the three measurement points. It was

---

[18] The SSS rating of one LDW driver was not available for the practice session.

tested whether there was a difference in quadratic trend for the two subgroups of HC drivers. This was found to be the case, $F(1, 13) = 5.18$, $p = .04$ (two-tailed), $\eta^2_G = .02$.

An ANOVA was conducted in order to see whether there was a significant main effect of the level of lane keeping assistance. Indeed, there was, $F(3, 39) = 5.34$, $p = .004$, $\eta^2_G = .21$[19]. Post-hoc comparisons with Bonferroni-corrected α-levels indicated that only the difference between NA drivers and drivers in the 'HC long' subgroup, $F(1, 39) = 10.03$, $p = .003$, $r = .45$, and the difference between NA drivers and drivers in the 'HC short' subgroup, $F(1, 39) = 10.94$, $p = .002$, $r = .47$, were significant.

4.4.3.1 Influence of drivers' energetic arousal on their reliance

Correlation analyses were conducted in order to explore whether drivers' reliance on the lane keeping assistance systems (as measured by the mean duration of their single glances away from the road scene) was related to their energetic arousal. For this purpose, HC and LDW drivers' ratings on the SSS were correlated with their mean duration of single display glances in Session 1 and in Session 2.

There was a positive correlation between HC drivers' ratings on the SSS and their reliance on the HC system, both for Session 1, $r(12) = .34$, $p = .12$ (one-tailed), and for Session 2, $r(12) = .30$, $p = .15$ (one-tailed)[20]. Note that higher scores on the SSS reflect lower levels of energetic arousal. Thus, a lower activation of HC drivers was related to a higher reliance on the HC system. The correlations between drivers' energetic arousal and their reliance on the HC system were however weaker than the correlations found for drivers' trust in the HC system and their reliance on it.

Drivers' reliance on the LDW system was found to be almost not related to their energetic arousal. There was a weak correlation in the opposite direction between LDW drivers' ratings on the SSS and their reliance on the LDW system in Session 1, $r(9) = -.13$, $p = .71$ (two-tailed), and a weak positive correlation for Session 2, $r(9) = .12$, $p = .28$ (one-tailed)[21].

---

[19] One NA driver was identified as outlier (with values laying more than 2.5 *SD* above the group means) and therefore excluded from the analysis.

[20] One HC driver was identified as outlier in the scatter plots and therefore excluded from the analysis.

[21] The reliance data of three LDW drivers was not available due to insufficient gaze tracking quality. One LDW driver was identified as outlier in the scatter plots and therefore excluded from the analysis.

4.4.3.2 Summary and discussion

It was assumed that drivers' reliance on the lane keeping assistance systems would not only be influenced by motivational processes (i.e., drivers' strategic investment of mental effort) and drivers' attitudes towards the system (i.e., their trust in it), but also by energetic processes. It was hypothesised that drivers' reliance on the lane keeping assistance systems increases as their energetic arousal and their ability for vigilant system monitoring decreases. It was further assumed that the increasing automation of the lateral control task and the herewith associated reduced manual control demands would lead to a decrease in drivers' activation. Thus, HC drivers were predicted to report the lowest level of arousal during the practice session and in Session 1, whereas NA drivers were predicted to report the highest level of arousal.

The results were fairly consistent with this hypothesis. There was a significant main effect of the level of lane keeping assistance on drivers' arousal as measured by drivers' ratings on the Stanford Sleepiness Scale (SSS). HC drivers reported a significant lower level of arousal than NA drivers. The effect of the level of lane keeping assistance was most pronounced during the 10-min practice session where drivers drove without concurrently performing the secondary task.

There was also some evidence for the predicted change in drivers' arousal over the course of the study as a function of the level of lane keeping assistance. According to the hypotheses, drivers' activation levels converged over the course of the study. Thus, NA drivers reported a significant decrease in arousal as a response to the increase in task demands from the practice session over Session 1 (resulting from the change from single-task to dual-task conditions) to Session 2 (resulting from additional occurrence of critical driving situations). It was assumed that this continuous decrease in NA drivers' arousal results from a depletion of attentional resources over time.

HC drivers differed significantly in their activation pattern. Whereas drivers in the 'HC short' subgroup who were found to rely not at all on the HC system showed an increase in arousal over the course of the study, drivers in the 'HC long' subgroup (who were found to rely to a large extent on the HC system) reported a decrease in arousal from the practice session to Session 1, and an increase in arousal from Session 1 to Session 2 approaching the activation level of the 'HC short' subgroup.

It was assumed that the differences in slopes in drivers' arousal over the course of the study as a function of the level of lane keeping assistance would result from the mental effort regulation strategies that drivers had adopted during the first half of the study. NA drivers were assumed to have invested the highest mental effort in the driving task in Session 1 because they could no rely on a lane keeping assistance system. A further increase in task demands caused by the occurrence of critical driving situations in Session 2 was assumed to lead to a further depletion of attentional resources that in turn should lead to a decline in arousal. On the contrary, HC drivers were expected to adopt to a larger extent an effort conservation strategy during Session 1. Thus, HC drivers were assumed to react to the increase in task demands in Session 2 by a mobilisation of 'reserve' attentional resources which should lead to an increase in arousal.

The results on drivers' mental effort regulation strategies however gave no clear indication for the predicted differences in drivers' investment of mental effort as a function of the level of lane keeping assistance in Session 1. Although there were no significant differences between groups, HC drivers (especially drivers in the 'HC long' subgroup) reported a higher level of overall mental workload than NA drivers in Session 1. Also, there was no clear indication that HC drivers had invested substantially less mental effort in the driving task than NA drivers in Session 1. Thus, the results of this study do not provide clear support for the assumption that drivers' changes in activation over the course of the study can be attributed to their differences in mental effort regulation strategies as a function of the level of lane keeping assistance. As drivers' arousal was not measured before the practice session, it cannot be completely ruled out that drivers already differed in their level of activation before they started to drive in the simulator. As participants were assigned randomly to the three experimental groups and the sequence of experimental conditions was matched over the day, it appears however unlikely that a-priori group differences in arousal may have caused the large observed effects.

The results on drivers' arousal also indicate that there seems to be no direct inverse relationship between drivers' level of activation and their reliance on the lane keeping assistance systems. Drivers in the 'HC short' subgroup who had been found to rely not at all on the HC system and who were less prepared than NA drivers to allocate their visual attention away from the road scene did not report the highest level of activation, but a significantly lower arousal than NA drivers. Furthermore, the two subgroups of HC drivers who had been found to differ significantly in their reliance on the HC system reported an equally low level of arousal, although they differed in their arousal pattern over the course of the study. LDW drivers and NA drivers who had been found to not differ in their preparedness to allocate their visual attention away from the road scene (as a measure of reliance) differed, although not significantly, in their subjective arousal.

The correlation analyses revealed that drivers' arousal played a role in drivers' reliance on the HC system, but apparently not in drivers' reliance on the LDW system. There was a medium correlation ($r \geq .3$) between HC drivers' scores on the SSS and their reliance on the HC system as measured by the mean duration of their single glances away from the road scene. Thus, the lower HC drivers' arousal, the higher was their reliance on the HC system. The correlation was however weaker than the correlation between HC drivers' trust in the HC system and their reliance on it, suggesting that trust had a more important influence on drivers' reliance on the HC system than drivers' level of activation.

To further explore the relative influences of drivers' attitudes towards the lane keeping assistance systems (i.e., their trust in it), of motivational processes (i.e., their strategic investment of mental effort), and of energetic processes (i.e., their arousal) on drivers' reliance on lane keeping assistance systems, multiple regression analyses were performed.

4.4.4 The relative influence of drivers' attitudes and of energetic and motivational processes on their reliance

A major assumption in this study was that drivers' reliance on a lane keeping assistance system is influenced by drivers' attitudes towards the system (i.e., their trust in it), as well as by energetic processes (i.e., arousal and vigilance) and motivational factors (i.e., drivers' strategic investment of mental effort). Correlation analyses had revealed that drivers' trust in the system and their subjective arousal were both related to drivers' reliance on the HC system, but not on the LDW system. These results let suggest that different processes may play a role in determining drivers' reliance on a lane keeping assistance system depending on the level of automation of the lateral control task. In order to further explore this hypothesis and to determine the relative influences of drivers' trust in the lane keeping assistance systems, their arousal, and their mental effort regulation strategies on their reliance on the HC and on the LDW system, multiple regression analyses were performed.

Hierarchical multiple regression was used to determine those factors which served as the best predictors for drivers' reliance. The regressions were performed based on drivers' data in Session 1 and in Session 2.

4.4.4.1 Factor analysis of drivers' mental effort regulation strategies

Drivers' mental effort regulation strategies were assessed by multiple items asking for the relative importance of different goals that drivers were assumed to pursuit in their strategic management of mental effort. More specifically, those items referred to the relative importance that drivers had assigned to the driving task vs. to the Arrows Task (secondary task) when they drove with different levels of lane keeping assistance and concurrently performed the Arrows Task. Only one item asked directly for drivers' prioritisation of the driving task vs. the Arrows Task by asking them to indicate the proportion of their overall mental effort they had invested in the driving task vs. in the Arrows Task during the preceding drive. It can be argued however that this item may have been insensitive to the absolute magnitude of mental effort that drivers had invested in the driving task and in the Arrows Task, because a reduced effort investment in the driving task may not automatically imply an increase in effort investment in the Arrows Task. For this purpose it was explored whether the remaining items asking for drivers' mental effort regulation strategies would reveal a meaningful factorial structure when entered into a factor analysis.

The factor analysis was based on drivers' ratings in Session 1 and in Session 2. Initially, five items were entered into the factor analysis (principal component analysis with Varimax rotation): two items referring to drivers' motivation to perform well on the Arrows Task, two items referring to drivers' motivation to protect their driving performance from deteriorations due to their concurrent performance of the Arrows Task, and one item referring to drivers' motivation to limit their effort investment in the driving task. The factor analysis yielded a two-factor solution. It became however apparent the variance in the last item ("I tried to maintain an acceptable level of driving performance without investing too much effort in it.") could not be sufficiently explained by the two factors (R^2 = .14). This item was therefore removed from the analysis. A subsequent factor analysis with the four remaining items yielded again a two-factor solution which accounted for 70.67% of the total variance. The factor analysis grouped the two items referring to drivers' motivation to perform well on the Arrows Task in one factor, and the two items referring to drivers' motivation to protect their driving performance from deteriorations due to their concurrent performance of the Arrows Task in the second factor. The fact that the four items did not load on the same factor supports the notion that drivers' attempts to maintain acceptable driving performance standards and their attempts to perform well on the Arrows Task were two distinct and probably concurrent motives influencing their mental effort regulation strategies. The first factor will be subsequently labelled as "motivation to perform well on the Arrows Task" and the second factor will be labelled as "motivation to protect driving performance". The factor loadings of the four items are presented in Table 18.

Table 18. Factor pattern loadings and communality of mental effort regulation items for the rotated solution ($N = 90$).

Item	Factor 1 (Perform well on the Arrows Task)	Factor 2 (Protect driving performance)	Communality
I wanted to succeed very much on the Arrows Task.	-.85	-.13	.75
I didn't take the Arrows Task too seriously.	.87	.01	.76
It was important for me that the Arrows Task did not have a negative effect on my driving performance.	-.06	.83	.70
I could not perform as well as I wanted on the Arrows task. The driving task required all of my attention.	.20	.76	.62

4.4.4.2 Multiple regression analyses

Drivers' factor scores on the two mental effort regulation factors, their trust ratings, and their scores on the SSS in both sessions were used as predictors of their reliance on the lane keeping assistance systems as measured by the mean duration of their single glances away from the road scene in both sessions. Thus, for each driver two sets of variables were entered into the regression analysis, one for Session 1 and one for Session 2.

Determinants of drivers' reliance on the HC system

Hierarchical multiple regression analyses in which different combinations of the above mentioned predictors were tested revealed that drivers' trust in the HC system together with HC drivers' arousal served as the best predictors for drivers' reliance on the HC system, $F(2, 25) = 9.16$, $p = .001$[22]. Trust and arousal together accounted for 42.3% of the variance in drivers' reliance on the HC system. Drivers' trust in the HC system had a slightly greater influence on their reliance on the HC system, $\beta = .58$, $t(25) = 3.60$, $p = .001$, than drivers' arousal, $\beta = .53$, $t(25) = 3.33$, $p = .003$. Note that higher scores on the SSS reflect lower levels of arousal. Thus, the higher drivers' trust in the HC system and the lower their arousal, the more did they rely on the HC system. The two mental effort regulation factors did neither significantly improve the predictive power of the model, nor did they explain a sufficient variance in drivers' reliance themselves when entered as sole predictors in the model.

Determinants of drivers' reliance on the LDW system

LDW drivers' motivation to perform well on the Arrows Task turned out to be the single best predictor of drivers' reliance on the LDW system, $F(1, 20) = 7.45$, $p = .01$, explaining 27.1% of the variance in drivers' reliance. Thus, overall, drivers' reliance on the LDW system could be less well predicted by the three major assumed underlying factors than drivers' reliance on the HC system. The higher LDW drivers' motivation to perform well on the Arrows Task, the more did they rely on the LDW system, $\beta = -.52$, $t(20) = -2.73$. Note that high factor scores reflected drivers' low motivation to perform well on the Arrows Task (see factor loadings in Table 18). Neither drivers' motivation to protect their driving performance, nor drivers' arousal or their trust in the LDW system explained a significant proportion of the variance in drivers' reliance on the LDW system.

4.5 Drivers' Evaluation of the Lane keeping Assistance Systems

Drivers were asked to evaluate different aspects of their interaction with the lane keeping assistance systems in order to yield a more complete picture about their adaptation to the increasing automation of the lateral control task. Drivers were asked to evaluate the HC and the LDW system in terms of their interventions in driving, their usability, controllability, and their supposed effects on driving behaviour and safety. The results concerning these different evaluation dimensions will be reviewed in turn.

---

[22] One HC driver was identified as outlier in the residual analysis and therefore excluded from the analysis.

4.5.1 Adequacy of the lane keeping assistance systems' interventions in driving

Drivers' were asked to evaluate the adequacy of the HC and the LDW systems' interventions in driving in terms of the frequency of interventions, the time of intervention, and the strength of intervention. The results indicate that both systems' interventions in driving were perceived by a majority of the drivers as appropriate (see Table 19).

Table 19. Drivers' evaluation of the adequacy of the HC system's and the LDW system's interventions in driving in Session 1 and Session 2. Presented are the numbers of drivers per evaluation category.

	HC		LDW	
Evaluation dimension	Session 1	Session 2	Session 1	Session 2
Intervention frequency				
too often	0	3	0	1
appropriate	11	10	12	12
too rarely	1	1	0	1
Intervention strength				
too strong	1	1	0	0
appropriate	9	11	10	11
too weak	2	2	2	3
Intervention time				
too early	0	1	1	1
appropriate	12	13	10	12
too late	0	0	1	1
n[a]	12	14	12	14

[a] Three HC drivers and three LDW drivers reported that the system had not intervened in Session 1. One HC driver and one LDW driver reported that the system had not intervened in Session 2.

4.5.2 System usability and controllability

Drivers were asked to evaluate their interaction with the HC and the LDW system in terms of usability and controllability. Usability items referred to drivers' joy of using the systems, the systems' learnability and to the monitoring demands the systems placed on drivers. Controllability items referred to drivers' ability to identify and to react appropriately to critical driving situations representing functional system limits. All items were 7-point ratings scales with the endpoints "strongly disagree" and "strongly agree".

Figure 49 presents drivers' evaluation of the lane keeping assistance systems in terms of the three aspects of system usability (joy of use, learnability and monitoring demands, identified as (a), (b), and (c), respectively in Figure 49) and in terms of controllability (identified as (d) in Figure 49).

There were no considerable differences between the two subgroups of HC drivers (with $r \geq |.3|$, corresponding to a medium effect size) for any but one item referring to the HC system's controllability. Thus, drivers' ratings on this item ("I could easily identify situations where I had to take over control over the system.") are presented separately for drivers in the 'HC long' subgroup (who had been found to rely to a high extent on the HC system) and for drivers in the 'HC short' subgroup (who had been found to not rely on the HC system at all).

As Figure 49 shows, drivers' evaluations of the lane keeping assistance systems were generally quite positive. Independent t-tests were conducted for each item in order to find out whether drivers differed significantly in their evaluations of the HC and the LDW system.

The results of these statistical analyses along with the group means and standard deviations are listed in Table 20. HC drivers and LDW drivers were found to not differ in their evaluations of the systems' monitoring demands and controllability; however they differed significantly in their joy of driving with both systems, and in their ratings of the HC and the LDW systems' learnability. Thus, drivers reported that they enjoyed it to a significantly higher degree to drive with the LDW system than to drive with the HC system. Furthermore, they perceived driving with the LDW system to be significantly easier than driving with HC system, and reported that they needed significantly more time to learn how the HC system worked than to learn how the LDW system worked (see rows 1 up to 3 in Table 20).

An independent t-test contrasting the two subgroups of HC drivers in their mean rating on the above mentioned controllability item revealed that the difference between drivers in the 'HC long' subgroup and drivers in the 'HC short' subgroup was only marginally significant, $t(13) = -1.98$, $p = .07$ (two-tailed), although the effect was quite large, $r = -.48$. Thus, drivers in the 'HC short' subgroup reported that it was easier for them than for drivers in the 'HC long' subgroup to identify situations where they had to take over control over the HC system. Because the difference between drivers in the 'HC long' subgroup and LDW drivers was smaller, as indicated in Figure 49, a one-way ANOVA with the three groups of drivers (HC long, HC short, LDW) as levels of the between-subjects factor was considered to not reveal a significant effect, and was therefore omitted. The two HC subgroups did not differ in their ratings of any other controllability item.

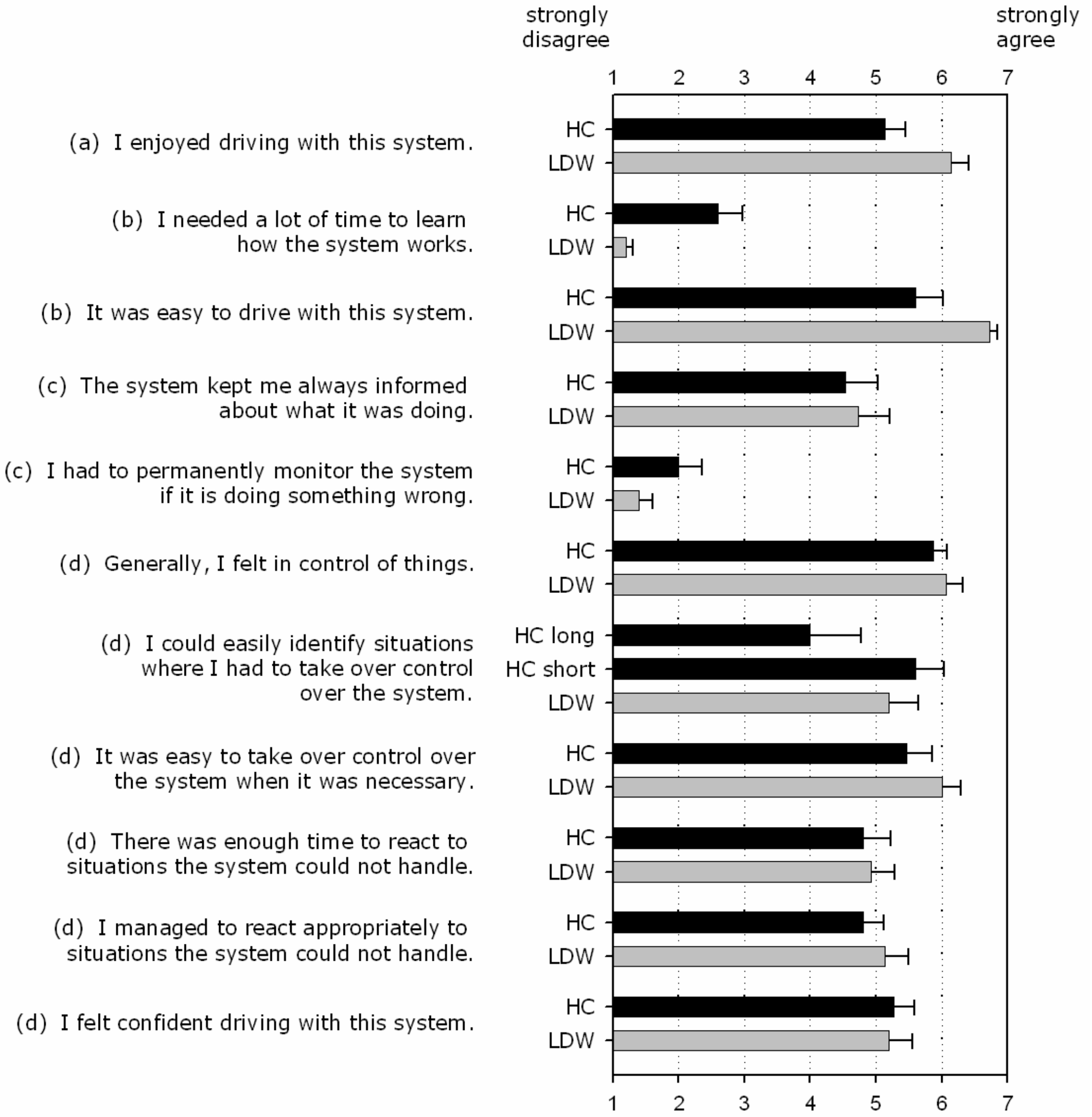


Figure 49. Drivers' evaluation of their interaction with the HC system and the LDW system in terms of joy of use (a), learnability (b), monitoring demands (c), and controllability (d). Drivers differed significantly in their ratings of the systems' learnability (b), and in their joy of driving with the systems (a). Drivers in the 'HC long' subgroup and drivers in the 'HC short' subgroup are represented by separate bars when they were found to differ considerably ($r \geq .3$) in their evaluation of their interaction with the HC system.

Table 20. Drivers' evaluation of the lane keeping assistance systems in terms of usability and controllability. Presented are group means and standard deviations for HC and LDW drivers and results of independent t-tests of the difference between HC drivers and LDW drivers for each item.

Item	Group	*M*	*SD*	*t*(28)	p^{a}	*r*
I enjoyed driving with this system.	HC	5.13	1.25	[b]-2.37*	.03	-.42
	LDW	6.14	1.03			
I needed a lot of time to learn how the system worked.	HC	2.60	1.45	3.59**	.002[c]	.56
	LDW	1.20	.41			
It was easy to drive with this system.	HC	5.60	1.64	-2.58*	.02[d]	-.44
	LDW	6.73	.46			
The system kept me always informed about what it was doing.	HC	4.53	1.92	-.29	.77	-.06
	LDW	4.73	1.83			
I had to permanently monitor the system if it is doing something wrong.	HC	2.00	1.36	1.46	.16	.27
	LDW	1.40	.83			
Generally, I felt in control of things.	HC	5.87	.83	-.61	.55	-.11
	LDW	6.07	.96			
I could easily identify situations where I had to take over control over the system.	HC long[e]	4.00	1.73			
	HC short[f]	5.60	1.35			
	LDW	5.20	1.70			
It was easy to take over control over the system when it was necessary.	HC	5.47	1.51	-1.10	.28	-.20
	LDW	6.00	1.13			
There was enough time to react to situations the system could not handle.	HC	4.80	1.61	-.24	.81	-.05
	LDW	4.93	1.39			
I managed to react appropriately to situations the system could not handle.	HC	4.80	1.21	-.70	.49	-.13
	LDW	5.13	1.41			
I felt confident driving with this system.	HC	5.27	1.22	.14	.89	.03
	LDW	5.20	1.37			

[a]two-tailed. [b]*df* = 27 (the data of one LDW driver was considered as outlier and therefore excluded from the analysis). [c]*df* = 16.255 (adjusted for unequal variances among groups). [d]*df* = 16.171 (adjusted for unequal variances among groups). [e]*n* = 5. [f]*n* = 10.

$^{*}p < .05$. $^{**}p < .01$.

4.5.3 Supposed effects of the lane keeping assistance systems on driver behaviour and safety

Drivers were also asked to evaluate the HC system and the LDW system in terms of their supposed effects on driver behaviour and safety. The results are depicted in Figure 50. The two HC subgroups are represented by separate bars when they were found to differ considerably ($r \geq |.3|$, corresponding to a medium effect size) in their ratings.

As Figure 50 reveals, drivers thought that the HC system and the LDW system would to a greater extent increase drivers safety than driver comfort (see the first two items on top of Figure 50). A further look at Figure 50 indicates that drivers did not envisage any pronounced negative effects of the two lane keeping assistance systems on driver behaviour and safety. However, there were some differences between groups. Independent t-tests were conducted to contrast HC drivers and LDW drivers when the two subgroups of HC drivers were found to not differ markedly in their ratings ($r < |.3|$). Results of these statistical tests, along with the group means and standard deviations are presented in Table 21.

HC and LDW drivers were found to differ significantly in their ratings of the systems' induced monotony. Drivers thought that driving with the HC system would increase monotony while driving to a significant higher extent than driving with the LDW system.

Drivers also differed significantly in their ratings of the extent to which the lane keeping assistance systems would encourage them to pay more attention to things unrelated to driving. A one-way ANOVA with the two HC subgroups and LDW drivers as levels of the between-subjects factor revealed a significant effect of group, $W(2, 16.58) = 5.42$, $p = .02$ (using Welch's statistic in order to correct for unequal variances among groups). Post-hoc Games-Howell tests revealed that only the difference between LDW drivers and drivers in the 'HC long' subgroup was statistically significant. Thus, drivers in the 'HC long' subgroup who had been found to rely to a high extent on the HC system thought that the system would encourage them to a significantly greater extent to pay more attention to things unrelated to driving compared to LDW drivers. Drivers in the 'HC long' subgroup felt also more encouraged by the HC system to pay more attention to things unrelated to driving compared to drivers in the 'HC short' subgroup (who had been found to not rely on the HC system at all), $t(12) = 1.98$, $p = .07$ (two-tailed), $r = .38$ (*df* adjusted for unequal variances among groups).

The two HC subgroups differed also in their perception of the degree to which the HC system would help them to focus their attention to the road, and of the degree to which the HC system would encourage them to drive riskier. Drivers in the 'HC short' subgroup thought that the HC system would help them to a greater extent to focus their attention to the road than drivers in the 'HC long' subgroup, $t(12.59) = -1.88$, $p = .08$ (two-tailed), $r = -.40$ (adjusted for unequal variances among groups). Drivers in the 'HC long' group reported that the HC system would encourage them to a greater extent to drive riskier compared to drivers in the 'HC short' subgroup, *t*(13) = 1.32, *p* = .21 (two-tailed), *r* = .35. One-way ANOVAs (comparing all three groups of drivers) were not conducted for these two items because the difference between the most extreme groups (the two HC subgroups) had already turned out to be not significant.

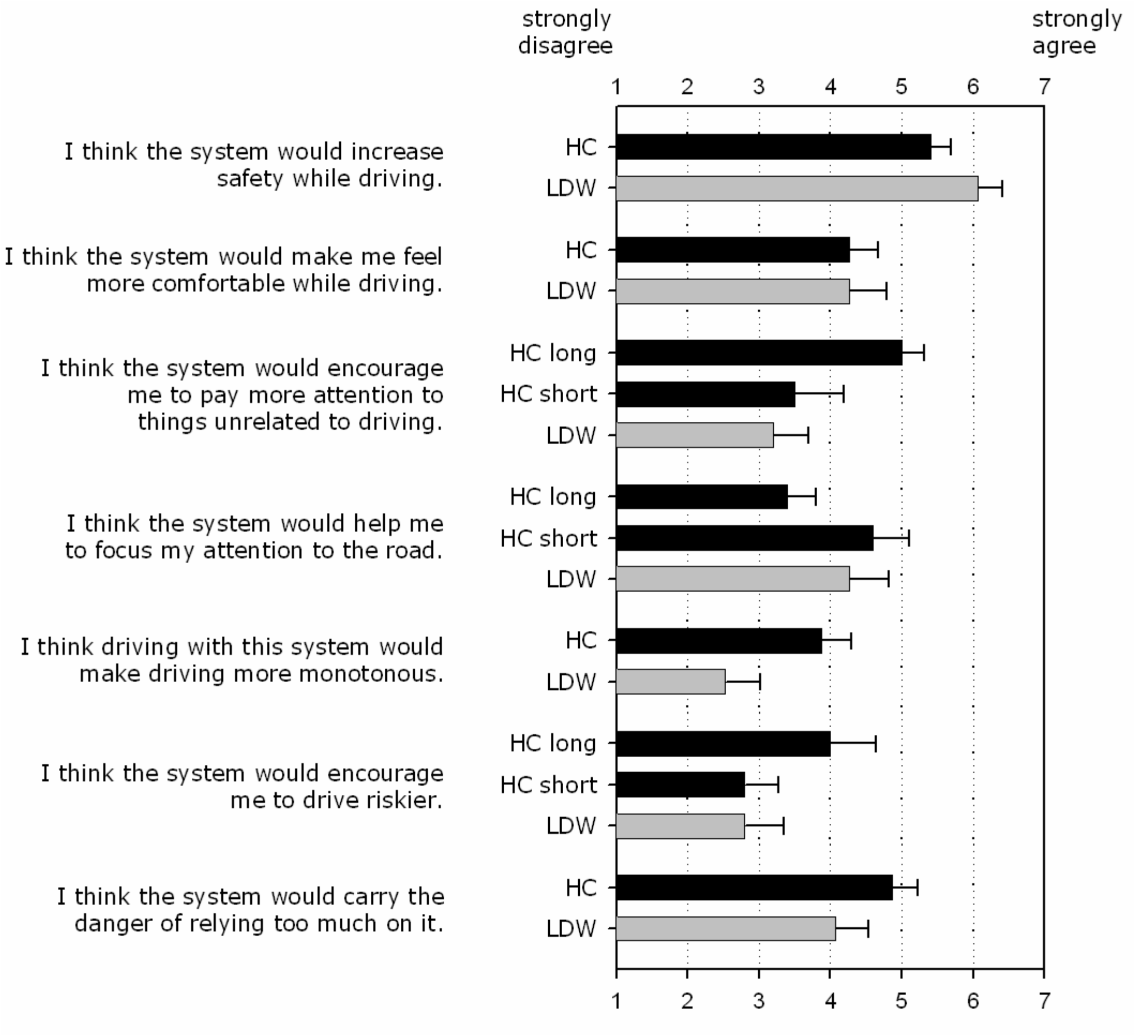


Figure 50. Drivers' evaluation of the HC system and the LDW system in terms of their supposed effects on driver behaviour and safety. Drivers in the 'HC long' subgroup and drivers in the 'HC short' subgroup are represented by separate bars when they were found to differ considerably (*r* ≥ .3) in their evaluation of the HC system.

Table 21. Drivers' evaluation of the lane keeping assistance systems in terms of their supposed effects on driver behaviour and safety. Presented are group means and standard deviations for HC and LDW drivers and results of independent t-tests of the difference between HC drivers and LDW drivers for items where the two subgroups of HC drivers did not differ considerably in their ratings ($r < .3$).

Item	Group	*M*	*SD*	*t*(28)	p^a	*r*
I think the system would increase safety while driving.	HC	5.40	1.12	-1.51[b]	.14	-.28
	LDW	6.07	1.27			
I think the system would make me feel more comfortable while driving.	HC	4.27	1.53	.00	1.00	.00
	LDW	4.27	2.02			
I think the system would encourage me to pay more attention to things unrelated to driving.	HC long	5.00	.71			
	HC short	3.50	2.17			
	LDW	3.20	1.90			
I think the system would help me to focus my attention to the road.	HC long	3.40	.89			
	HC short	4.60	1.58			
	LDW	4.27	2.12			
I think driving with this system would make driving more monotonous.	HC	3.87	1.64	2.07*	.05	.36
	LDW	2.53	1.89			
I think the system would encourage me to drive riskier.	HC long	4.00	1.41			
	HC short	2.80	1.75			
	LDW	2.80	1.82			
I think the system would carry the danger of relying too much on it.	HC	4.87	1.36	1.38	.18	.25
	LDW	4.07	1.79			

[a]two-tailed. [b]$df = 27$ (the data of one LDW driver was as outlier excluded from the analysis).
*$p < .05$.

4.6 Drivers' Situation Awareness

Situation Awareness (SA) was the second major concept (besides reliance) that was referred to in order to study changes in drivers' active engagement in the driving task as a function of the increasing automation of the lateral control subtask.

Drivers' SA was assessed by objective measures of drivers' reaction to different types of critical driving situations that occurred unexpectedly during Session 2, as well as by subjective measures.

Objective measures of drivers' SA were implicit performance measures of the latency and magnitude of drivers' reaction to different critical driving situations. More specifically, two types of objective measures were chosen as indicators for drivers' SA: (a) measures that referred to the relative point in time of drivers' initial reaction to the critical driving situations (response latency), and (b) measures that referred to the abruptness of the driving manoeuvre that was initiated by drivers in response to their encountering of the situations (response magnitude). The basic assumption was that drivers' delayed and more abrupt reaction to critical driving situations would be indicative for an impairment of the cognitive processes necessary to instantiate (activate) specific environmental information within the mental representation of the actual situation (situation model).

The critical driving situations that occurred unexpected for drivers during Session 2 were designed to differ in the processing demands for visual information that was either relevant for the driving subtask under manual control (longitudinal control) or for the driving subtask that was partly under automated control (lateral control). Table 22 gives an overview about the different types of critical situations. Seven of the eight situations were "safety-critical" - meaning that a missing driver reaction in those situations could have potentially resulted in a collision with other traffic (see upper part of Table 22). The eighth situation was categorised as "not safety-critical". In this situation, drivers' tendency to follow the erroneous behaviour of a lead vehicle ("car-following", see lower part of Table 22) was used as a measure of their SA.

The *seven safety-critical* situations differed (a) to whether they required drivers to react primarily on the lateral or on the longitudinal control level in order to avoid a collision, (b) with respect to the "time horizon" available for perception of the relevant visual cues, response planning and execution, and (c) with respect to the demands they placed on higher-level cognitive processes, i.e. whether they required a high or a low level of anticipation.

Table 22. Types of critical driving situations designed to investigate changes in drivers' situation awareness when driving with different levels of lane keeping assistance. Situation categories were realised by specific driving “Events” (labelled alphabetically) that occurred unexpected for drivers in the second experimental session. Numbers in brackets refer to the number of critical situations occurring in the respective category. An effect of the level of lane keeping assistance was expected for Event C (lateral unforeseeable) and for Event G (lateral anticipatory).

Safety-critical (7)		
	Time horizon available	
Anticipation demands	small	large
high	--	“anticipatory” lateral *(Event G)* longitudinal *(Event F, Event E)*
low	“unforeseeable” lateral *(Event C)* longitudinal *(Event D)*	“foreseeable” lateral *(Event A)* longitudinal *(Event B)*
Not safety-critical (1)		
“car-following” longitudinal *(Event H)*		

It was hypothesised that with increasing automation of the lateral control subtask drivers allocate fewer attentional resources to the processing of visual cues relevant for lateral control of the vehicle. This reduced processing of information relevant for lateral control was assumed to result in failures in bottom-up control of attention in driving situations that required drivers to react by adjustments in lateral control of the vehicle (e.g., by drawing aside) in order to avoid a collision with other traffic. Thus, an effect of the level of lane keeping assistance was only expected in situations that required drivers to react on the lateral control level and not in situations requiring drivers to react primarily on the longitudinal control level. Moreover, it was assumed that drivers would be able to compensate for failures in bottom-up attentional control when they were left sufficient time to perceive and interpret environmental cues and to plan ahead their reaction. Thus, it was assumed that automation-induced impairments in drivers’ SA should particularly become apparent in situations that required drivers to react on the lateral control level *and* that did not leave drivers much time to perceive and interpret the relevant information and to prepare their reaction (in lateral unforeseeable situations – Event C). Furthermore, automation-induced impairments in drivers’ SA should become apparent in situations that required drivers to react on the lateral control level *and* that demanded for higher-level cognitive processing of the visual cues requiring drivers to anticipate the behaviour of other traffic participants (in lateral anticipatory situations – Event G).

Thus, an effect of the level of lane keeping assistance on the timeliness and magnitude of drivers' reaction (indicative for impairments in drivers' SA) was hypothesised to become apparent only in Event C and in Event G, with HC drivers showing the latest and most abrupt reactions to these situations, and NA drivers showing the earliest and least abrupt reactions to these situations. On the contrary, it was assumed that HC drivers, LDW drivers and NA drivers would not differ in the timeliness and abruptness of their reaction to the other six Events.

The following section explains how the behavioural measures used as indicators for drivers' SA were analysed. Afterwards, the results on drivers' reaction to the eight critical driving situations will be presented separately for each Event. At the end of this chapter, the results obtained for the subjective measures of drivers' SA will be reviewed in more detail.

4.6.1 Analysis of behavioural measures used as indicators for drivers' SA

As mentioned above, two types of behavioural measures were used as indicators for drivers' SA: (a) Measures that referred to the relative point in time of drivers' initial reaction to the critical driving situations and (b) measures that referred to the abruptness of the driving manoeuvre that was initiated by drivers in response to their encountering of the situations.

In order to yield a more general picture of the effects of the level of lane keeping assistance on drivers' behaviour in safety-critical driving situations, some other measures of driving performance were assessed for explorative purposes. Those measures were however not taken as indicators for drivers' SA.

Table 23 lists some standard measures that were individually analysed for each driver for each of the safety-critical situations[23]. The different nature of the critical driving situations sometimes required the analysis of other or additional measures of driving performance. Because drivers were found to react always (despite in some exceptional cases) on the longitudinal control level to the Events, and because this reaction on the longitudinal control level mostly represented drivers' first reaction to the Events; the abruptness of drivers' reactions was assessed by measures of longitudinal control only.

---

[23] In total, drivers encountered eight critical situations during the second experimental session, of which seven were „safety-critical" (requiring a reaction in order to avoid collisions with other traffic), and one was „not safety-critical". This latter situation (Event H) assessed drivers' SA by their tendency to follow the erroneous behaviour of a lead vehicle that broke a speed limit. The SA measure in this situation was drivers' reduction of speed (yes/no) when passing the speed limit, and *not* the timeliness of drivers' reaction or the magnitude of reaction (speed change).

Table 23. Standard driving performance measures of the timeliness and abruptness of drivers’ reaction to the critical driving situations used as indicators for drivers' SA when driving with different levels of lane keeping assistance. Measures that describe the general characteristics of the initiated driving manoeuvre (last row) were analysed for explorative purposes in order to investigate potential effects of the lane keeping assistance on drivers’ behaviour in critical driving situations. All measures were extracted individually for each driver for each of the critical driving situations.

Variable	Operational definition	Control level
Timeliness of drivers’ reaction to critical situation	Time to Collision (TTC)[a] [s] at the initial point of drivers’ reaction on either the longitudinal control level (release of accelerator pedal, braking) or on the lateral control level (drawing aside)	lateral & longitudinal
Abruptness of initiated driving manoeuvre	Time [s] between release of accelerator pedal and application of brake	longitudinal
	Maximum braking force [N]	longitudinal
Manoeuvre characteristics	Maximum lateral deviation [m] / Magnitude of change in lateral position [m]	lateral
	Minimum speed [km/h] / Magnitude of speed change [km/h]	longitudinal

[a] TTC was calculated as the time to drivers’ collision with situation-specific stationary or moving objects, e.g. oncoming vehicles, or vehicles occupying the drivers’ lane. In the case of stationary objects, Time Headway (THW) was used instead of TTC.

The first step in the analysis of these behavioural measures was to visually explore the behavioural profile of each driver for each of the eight critical situations. For this purpose, different plots were created for each driver that depicted their behavioural reaction on the lateral control level and on the longitudinal control level up from the distance where drivers were able to recognise the critical situations to about 300 m after they had passed the critical situations.

Afterwards, the different measures of the timeliness and abruptness of drivers’ reactions to the critical driving situations were extracted from the driver behaviour data recorded for each driver by the help of custom-written Matlab programs. Measures of the initial point of drivers’ reactions to the critical driving situations were verified by displaying them in the plots of drivers’ reactions on the longitudinal and on the lateral control level.

The timeliness of drivers’ reactions to the critical situations was operationally defined as the Time to Collision (TTC) or Time Headway (THW) at the starting point of drivers’ first observable reaction to the Events either on the lateral or on the longitudinal control level. TTC was defined as the time until drivers would have collided with other traffic when their speed was kept constant. If drivers' collision referred to stationary objects, THW was used instead of TTC.

The following section explains how the initial point of drivers' reaction on the lateral and on the longitudinal control level was determined.

4.6.1.1 Definition of the initial point of drivers' reaction on the longitudinal control level

Drivers' reactions on the longitudinal control level mostly referred to a reduction of speed when they approached the critical driving situations. In most cases, drivers' release of the accelerator pedal (AccPed) was the first observable reaction on the longitudinal control level as response to the occurring Events. The position of the accelerator pedal was measured on a range between 0 (complete release) and 1 (complete press) with a precision of three decimal places with a sampling rate of 200 Hz. In order to determine the starting point of drivers' release of the accelerator pedal, two successive moving time windows were defined that went back in intervals of 1/8 s up from the point where drivers had completely released the accelerator pedal (AccPed = 0) as long as the mean position of the accelerator pedal within the preceding 1/4 s time interval was at least .001 larger than the mean position of the accelerator pedal in the successive 1/4 s time interval.

The Matlab algorithm applied was:

```
z = first_Min_AccPed;
y = z - 50;
x = y - 50;

while (mean(data(x:y,AccPed)) - mean(data(y:z,AccPed)) ≥ 0.001)
          z = z - 25;
          y = z - 50;
          x = y - 50;
end

start_Release_AccPed = y;
```

where first_Min_AccPed was the first row in the data file where drivers had completely released the accelerator pedal; AccPed was the column in the data file where the position of the accelerator pedal was recorded; and x, y, and z were the three variables (rows in the data file) that defined the two successive 1/4 s moving time windows.

The above algorithm was able to detect the starting point of continuous, relatively fast releases of the accelerator pedal. Thus, by the help of the algorithm, definite reactions of drivers in terms of a release of the accelerator pedal could be detected. It has to be noted that the starting point detected by the algorithm did not have to necessarily coincide with the point where drivers had actually *perceived* the critical situations. Thus, it is possible that drivers had recognised the critical situations at an earlier point in time but did not immediately show a discernable reaction (by releasing the accelerator pedal). On the other hand, some drivers were found to show an ambiguous pattern in behaviour (e.g., a stepwise release of the accelerator pedal with sometimes intermittent acceleration periods) from which the actual starting point of their reaction was hard to determine. In those cases, the starting point of drivers' release of the accelerator pedal detected by the algorithm may have referred to a later point in time than the actual starting point of drivers' reaction.

If drivers did not press the accelerator pedal when they approached the Events, the point of their initial reaction on the longitudinal control level was defined as the point where they started to apply the brake. When drivers did neither release the accelerator pedal nor brake when approaching an Event, no reaction was recorded on the longitudinal control level. There were some instances where drivers reacted actually not by reducing their speed in response to an Event, but by accelerating - for example in order to overtake an obstacle on right road side. In these cases, the point where drivers began to (re-)press the accelerator pedal was defined as the initial point of their reaction to the Event on the longitudinal control level.

4.6.1.2 Definition of the initial point of drivers' reactions on the lateral control level

Drivers' reactions on the lateral control level referred to the drawing-aside or overtaking manoeuvres that drivers initiated in order to avoid collisions with other traffic. Two criteria were used to determine the starting point of these manoeuvres. The first criterion was based on the *Last Time to Line Crossing (TLC) Period*, and the second criterion was based on the *Standard Deviation (SD) of lateral Position*. The initial point of drivers' reaction on the lateral control level was then defined as the point in time where one of the two criteria applied first.

Definition of the criterion Last TLC Period

Time to line crossing (TLC) represents the time available for the driver until the moment at which any part of the vehicle reaches one of the lane boundaries if speed and steering wheel angle are kept constant (Godthelp, Milgram, & Blaauw, 1984). TLC was calculated with a procedure described in Östlund et al. (2005) which is based on an approximation of TLC using the first and the second derivative of lateral distance (Van Winsum, Brookhuis, & De Waard, 2000). The TLC measure yields typical waveforms which occur when the vehicle approaches one of the lane boundaries until it reaches a minimum TLC and than turns to move towards the opposite lane boundary due to a corrective steering action of the driver. Based on recommendations proposed in Östlund et al. (2004), TLC values larger than 20 s were ignored, as well as TLC periods with a duration of less than 1 s. Figure 51 depicts the TLC of one driver when approaching Event A. Event A was realised by a parked vehicle on the right road side partly occupying the drivers' lane that drivers had to overtake in order to avoid a collision. Positive TLC values refer to the driver's steering towards the left lane boundary, whereas negative TLC values refer to the driver's steering towards the right lane boundary. Zero crossings refer to a change in vehicle motion from the left boundary towards the right boundary, or vice versa. TLC is undefined (infinite) when the driver drives perfectly in the centre of the lane and the vehicle currently approaches any of the lane boundaries, whereas minimal lateral accelerations towards the lane boundaries result in very large values of TLC.

TLC is not defined when the front tyres of the vehicle are outside the lane boundaries.

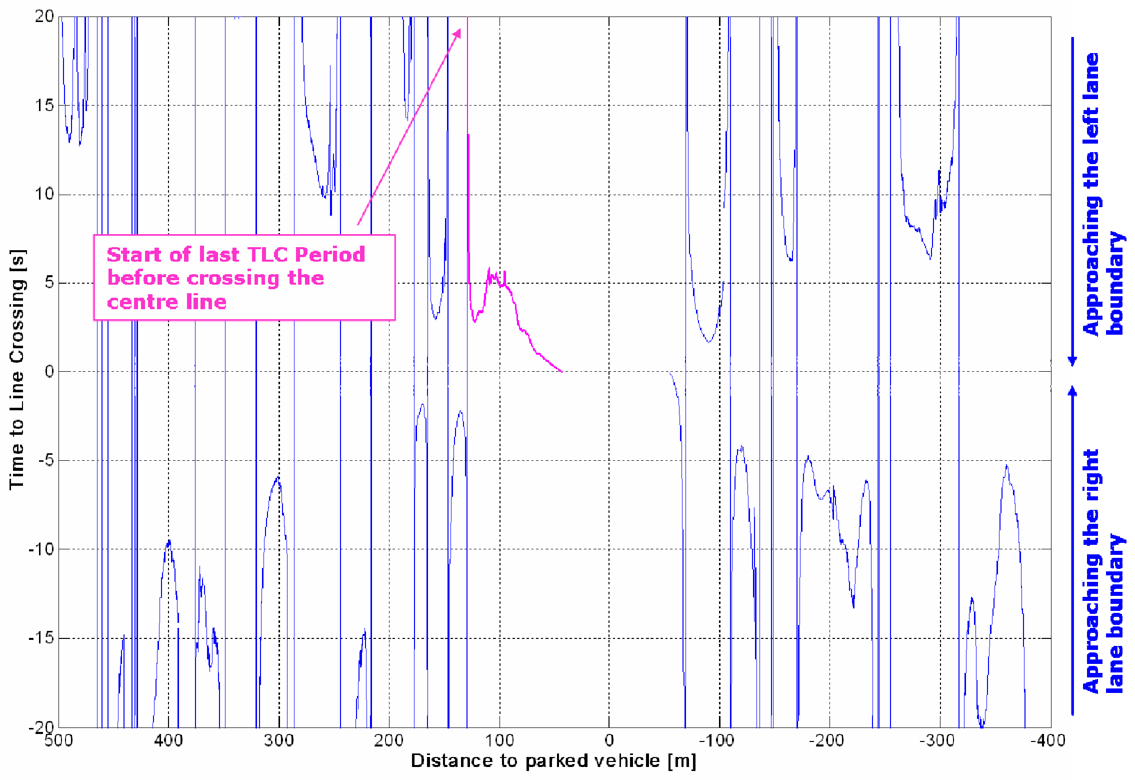


Figure 51. Time to Line Crossing [s] for one driver when approaching Event A (realised by a parking vehicle on the right road side partly occupying the drivers' lane). Lateral movements of the simulator vehicle towards the left lane boundary (centre line) are represented by positive TLC values, whereas lateral movements of the simulator vehicle towards the right lane boundary are represented by negative TLC values. TLC is not defined for periods where one of the front tyres is outside the lane boundaries. This is the case when the driver performed a lane change in order to overtake the parked vehicle (up from a distance of approximately 50 m before passing the vehicle to approximately 50 m afterwards in the figure). The starting point of the overtaking manoeuvre is defined as the starting point of the last TLC period before the vehicle crossed the centre line (marked in magenta).

The criterion *Last TLC Period* was defined as the starting point of the last TLC period (TLC < 20 s in case of avoidance manoeuvres to the left, or TLC > -20 s in case of avoidance manoeuvres to the right) before the front tyres of the vehicle crossed any of the lane boundaries. In case of only a near-by crossing of the lane boundaries, the criterion *Last TLC Period* was defined as the starting point of the TLC period with the smallest local minimum (in case of avoidance manoeuvres to the left) or the largest local maximum (in case of avoidance manoeuvres to the right) just before drivers passed the traffic which they risked colliding with.

The criterion *Last TLC Period* worked well when drivers initiated an avoidance manoeuvre by one "large" deflexion of the steering wheel in the respective direction. When drivers initiated the avoidance manoeuvre by numerous smaller subsequent steering inputs in the respective direction (e.g., when trying to override the steering torques of the HC system), the criterion *Last TLC Period* often resulted in a late detection of starting point of the manoeuvre. For this reason, a second criterion was applied, which is explained in the following section.

Definition of the criterion Standard Deviation (SD) of lateral Position

Based on the criterion *SD of lateral Position*, the initial point of drivers' reaction on the lateral control level was defined as a change in lateral position towards the respective lane boundary that was larger than the drivers' mean lateral position during the last 200 to 300 m before they encountered the critical situations plus/minus one mean standard deviation of lateral position. A similar approach was for example used by Gray and Regan (2000).

Both the mean lateral position as well as the mean standard deviation of lateral position was calculated for each driver individually. The mean standard deviation of lateral position was calculated over two up to three about 300 m-long route sections that were comparable in slope and curve to the road segments were the Events occurred. Furthermore, since drivers always performed the Arrows Task while driving when they encountered the critical driving situations, only route sections in Session 2 where drivers had concurrently performed the Arrows Task while driving were used for the calculation of the mean standard deviation of lateral position.

Figure 52 illustrates the basic measures of the timeliness and abruptness of drivers' reactions used as indicators for their SA when driving with different levels of lane keeping assistance.

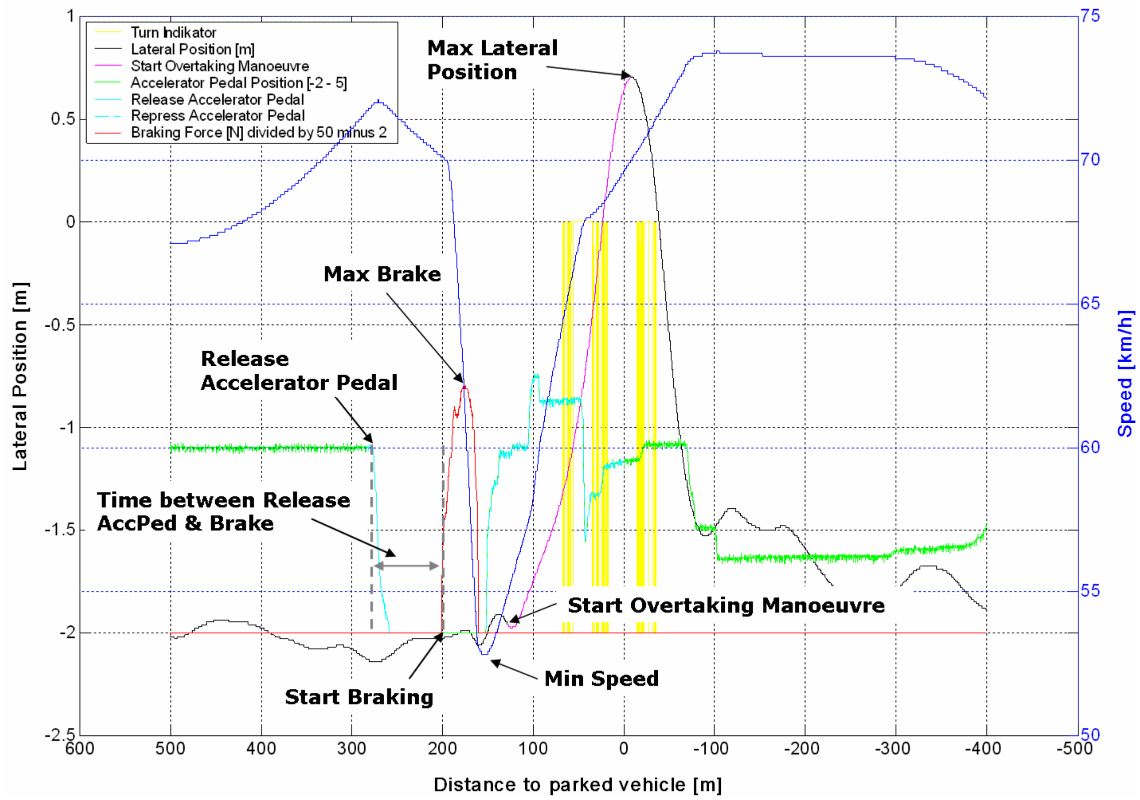


Figure 52. Illustration of the measures of the timeliness and abruptness of drivers' reactions to the critical driving situations used as indicators for their SA based on an example of one driver's reaction to Event A (realised by a parking vehicle on the right road side partly occupying the driver's lane). The driver first reacted by releasing the accelerator pedal (marked in cyan). After releasing the accelerator pedal, the driver braked and subsequently initiated the overtaking manoeuvre. The starting point of the overtaking manoeuvre (marked in magenta) is based on the criterion Last TLC Period which applied earlier than the criterion SD Lateral Position. The minimum speed and the maximum lateral position were used as measures of the general characteristics of the manoeuvre and not as indicators for drivers' SA (AccPed = Accelerator Pedal).

4.6.2 Drivers' reactions to foreseeable situations

4.6.2.1 Lateral foreseeable – Event A

Event A was realised by a parking vehicle on the right road side partly occupying the drivers' lane (see Figure 53) with its hazard warning lamp activated. The parking vehicle was visible for drivers approximately 300 m before.

Drivers had to realise that they would have to draw aside and to overtake the parking vehicle in order to avoid a collision. Thus, this Event required adjustments in lateral control of the vehicle and was designed to simulate functional limits of the HC and LDW system. Both systems offered inadequate support in this situation because they continued in assisting drivers to stay within the driving lane.

Event A was categorised as “foreseeable” because drivers had a comparably large time window to perceive and interpret the situation, and to plan ahead their response. Furthermore, the situation did not require drivers to anticipate the behaviour of other traffic participants and was unambiguous in terms of the manoeuvre that drivers were required to perform. Because of the low anticipation demands and the relatively large time horizon available for drivers to prepare their reaction, the hypothesised automation-induced failures in bottom-up control of attention were expected to not become apparent in drivers’ reaction to Event A. Thus, it was assumed that HC drivers, LDW drivers and NA drivers would not differ in terms of the timeliness and abruptness of their reaction to Event A, used as indicators for their SA.

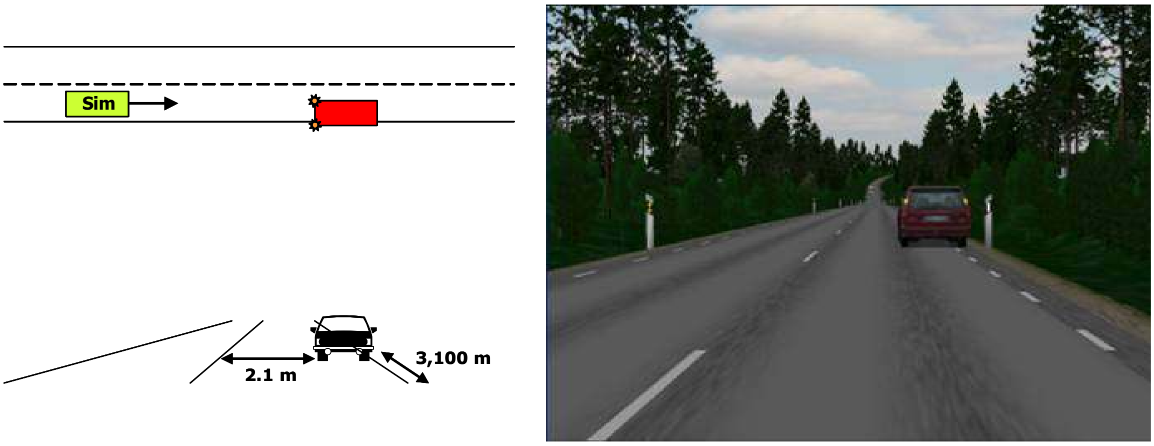


Figure 53. Pictorial view (left) and simulation screen-shot (right) of Event A from a bird's eye of view (top left) and from the driver's perspective (bottom left and right part of the figure). Drivers had to overtake the parking vehicle.

Manoeuvres initiated by drivers in response to Event A

All drivers reacted first on the longitudinal control level to Event A, either by releasing the accelerator pedal, by braking, or by accelerating in order to overtake the parking vehicle. The majority of drivers first released the accelerator pedal, then braked, and then initiated the overtaking manoeuvre, as can be seen in Figure 54.

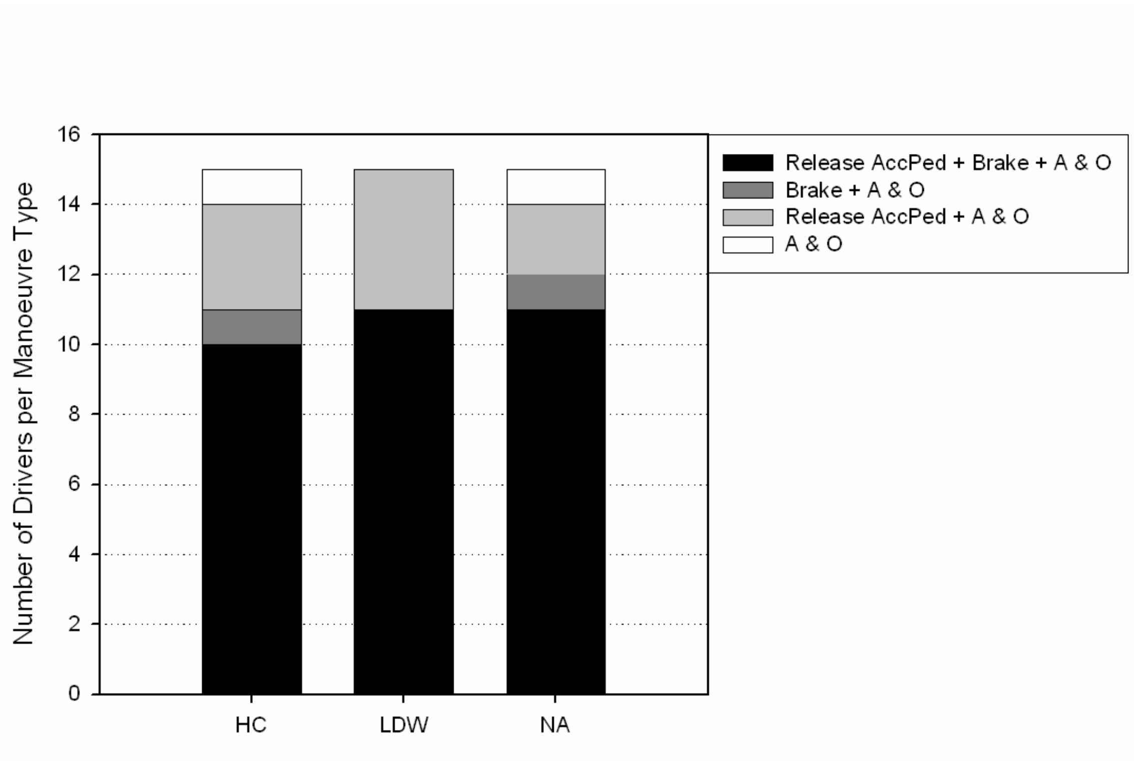


Figure 54. Types of manoeuvres initiated by drivers in reaction to Event A across the three levels of lane keeping assistance (AccPed = Accelerator Pedal, A & O = Accelerate and Overtake).

Timeliness of drivers' reaction to Event A

Drivers started to react on the average at a distance of 223.34 m (*SD* = 29.45) to the parking vehicle. In order to find out whether drivers differed in the latency of their reaction to Event A, the three levels of lane keeping assistance were compared in terms of the THW to the parking vehicle at the point where drivers started to react to Event A. The results are depicted in Figure 55.

As Figure 55 reveals, HC drivers (M = 11.39 s, SD = 2.31) reacted earlier than LDW drivers (M = 10.77 s, SD = 1.06) to Event A, whereas NA drivers (M = 10.28 s, SD = 1.40) reacted latest to Event A[24]. The differences between groups were not significant as revealed by a one-way ANOVA, $W(2, 25.97) = 1.33$, $p = .28$ (using Welch's statistic in order to correct for unequal variances among groups). The size of the effect was computed from the sum of squares of the ANOVA and yielded $\eta^2_G = .07$, which is a small effect.

---

[24] One LDW driver was excluded from the analysis because of an extreme value that lay more than 2.5 *SD* above the group mean.

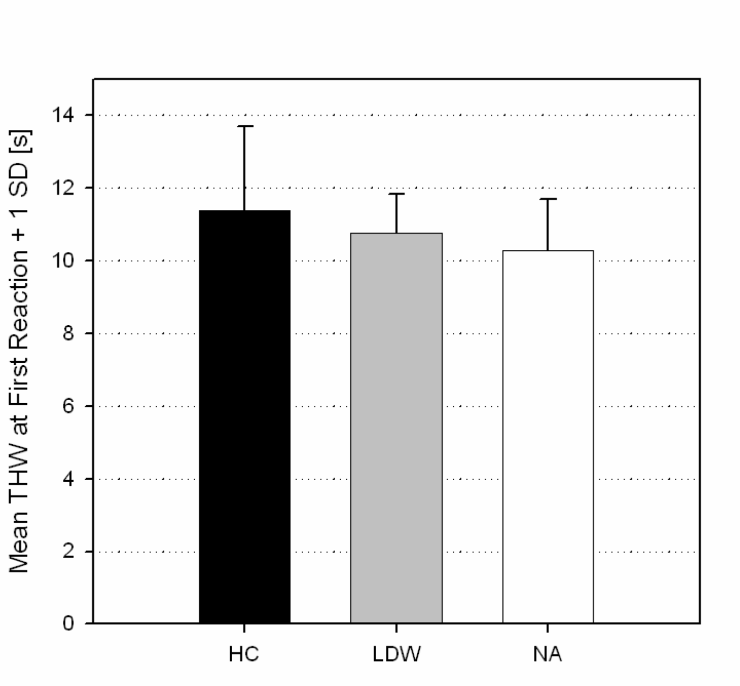


Figure 55. Mean Time Headway (THW) to the parking vehicle [s] at the initial point of drivers' reaction to Event A across the three levels of lane keeping assistance.

Abruptness of initiated driving manoeuvre

Two measures were used as indicators for the abruptness of the manoeuvre initiated by drivers: (a) the mean time between drivers' release of the accelerator pedal and their application of the brake, and (b) drivers' maximum braking force.

Time between drivers' release of the accelerator pedal and their application of the brake

LDW drivers applied the brake fastest after having started to release the accelerator pedal (M = 2.35 s, SD = 1.52), whereas HC drivers (M = 2.66 s, SD = 1.42) and NA drivers (M = 2.68 s, SD = 1.40) started to brake later after having started to release the accelerator pedal. A one-way ANOVA revealed no significant differences between groups, $F(2, 29) = .17$, $p = .84$, $\eta^2_G = .01$.

Maximum braking force

Drivers' maximum braking force was measured in N. HC drivers (M = 123.51, SD = 55.69) braked harder than NA drivers (M = 107.32, SD = 24.07) and LDW drivers (M = 100.47, SD = 35.76) in reaction to Event A[25]. A one-way ANOVA revealed that there were no significant differences between the three levels of lane keeping assistance, $W(2, 18.25) = .64$, $p = .54$ (using Welch's statistic in order to correct for

---

[25] One HC driver was excluded from the analysis because of an extreme value that lay more than 2.5 *SD* above the group mean.

unequal variances among groups). The size of the effect was computed from the sum of squares of the ANOVA and yielded $\eta^2_G = .06$, which is a small effect.

Influence of drivers' reliance on the lane keeping assistance systems on the timeliness and abruptness of their reaction to Event A

Researchers generally assume a negative relationship between operators' reliance on automation and their SA (Endsley, 1996; Sarter & Woods, 1991; Scerbo, 1996). It was tested whether drivers' reliance on the lane keeping assistance systems (measured by their preparedness to allocate their visual attention away from the road scene during performance of the visually demanding secondary task) was related to the timeliness and the abruptness of their reaction to Event A used as indicators for drivers' SA. For this purpose, measures of the timeliness and the abruptness of drivers' reactions to Event A were correlated with measures of drivers' visual attention allocation calculated over the 30-s Arrows Task period during which drivers encountered Event A. Two visual attention allocation measures were used: drivers' total glance time to the Arrows Task display up to the point where they started to react to Event A[26], and the mean duration of drivers' single glances to the Arrows Task display during Event A.

Indeed, a highly significant negative correlation was found between the THW at the initial point of drivers' reaction to Event A and drivers' total glance time to the Arrows Task display before they started to react to Event A, $r(35) = -.44$, $p = .006$. Thus, the longer drivers had (in sum) glanced to the Arrows Task display when they approached Event A, the later did they react to it. The THW at drivers' initial reaction was also significantly negatively correlated with the mean duration of drivers' single glances to the Arrows Task display during the Arrows Task where drivers encountered Event A, $r(35) = -.34$, $p = .04$.

Drivers' total glance time to the Arrows Task display when they approached Event A was also significantly negatively correlated with the time between drivers' release of the accelerator pedal and start braking, $r(25) = -.55$, $p = .003$. Thus, the longer drivers had (in sum) glanced to the Arrows Task display when they approached Event A, the more rapidly did they start to brake after they had begun to release the accelerator pedal. The correlation between drivers' single display glance duration during the Arrows Task where they encountered Event A and the time between drivers' release of the accelerator pedal and start braking was considerably weaker and not significant, $r(25) = -.15$, $p = .47$.

General characteristics of drivers' reaction to Event A

In order to reveal whether drivers in the three groups differed systematically in their reaction to Event A (which can be interpreted as an effect of the level of lane keeping assistance), some more general measures describing the characteristics of the manoeuvres initiated by drivers were analysed. For Event A, two measures were of

---

[26] The total glance time to the Arrows Task display was calculated only for drivers for whom the eye-tracking data was available for the whole duration of the Arrows Task.

interest: (a) the frequency of drivers' use of the turn indicator, and (b) drivers' lateral position when they were at one level with the parking vehicle on the right road side. Those measures did *not* serve as indicators for drivers' SA.

Drivers' use of the turn indicator

Drivers were required to overtake the parking vehicle on the right road side and thus, to override the steering torques of the HC system or to ignore the LDW system's warnings. Alternatively, HC drivers and LDW drivers could use the turn indicator to signal an intentional departure from the driving lane, which led to a temporary deactivation of the HC and the LDW system. As Figure 56 shows, drivers driving with a lane keeping assistance system used the turn indicator more often than drivers in the control group (NA drivers). This difference was however not significant, $p = .52$ (using Fisher's exact test), $\Phi = .21$.

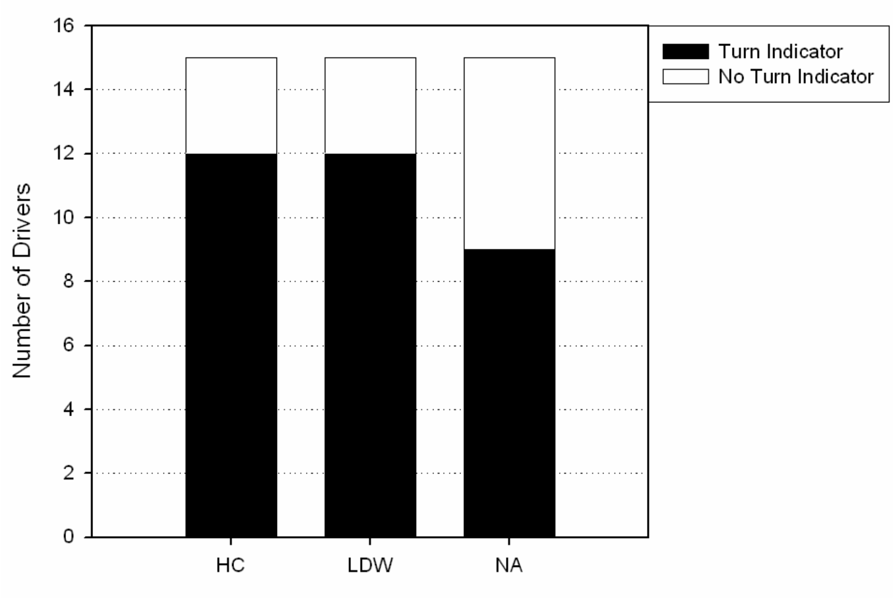


Figure 56. Drivers' use of the turn indicator during initiation of the overtaking manoeuvre in reaction to Event A.

Drivers' lateral distance to the parking vehicle

Drivers' lateral distance to the parking vehicle was measured as the distance between both vehicles at the point where both vehicles were on the same level. LDW drivers had the largest lateral distance to the parking vehicle ($M = 2.26$ m, $SD = 0.53$), followed by NA drivers ($M = 1.45$ m, $SD = 0.28$) and HC drivers ($M = 1.40$ m, $SD = 0.44$). A one-way ANOVA revealed no significant effect of the level of lane keeping assistance, $W(2, 25.89) = 1.16$, $p = .33$ (using Welch's statistic in order to correct for unequal variances among groups). The size of the effect was computed from the sum of squares of the ANOVA and yielded $\eta^2_G = .07$, which is a small effect.

4.6.2.2 Longitudinal foreseeable – Event B

Event B was (similarly to Event A) realised by a parking vehicle on the right road side partly occupying the drivers' lane with its hazard warning lamp activated. However, there were two oncoming vehicles that were programmed in way that drivers first had to let them pass before they could safely overtake the parking vehicle (see Figure 57). Thus, Event B required drivers to react first on the longitudinal control level. The oncoming vehicles approached the parking vehicle from the opposite direction with the same speed as the drivers (unless drivers' speed fell below 54 km/h, then the vehicles continued to drive with 54 km/h).

The parking car was visible for drivers approximately 360 m before. Event B was categorised as foreseeable because it was unambiguous and demanded from drivers a clear reaction on the longitudinal control level. Furthermore, drivers had a comparably large amount of time for situation perception and planning of an adequate response. Therefore it was predicted that the three groups of drivers would not differ in the timeliness and abruptness of their reaction (used as indicators for drivers' SA) as a function of the level of lane keeping assistance.

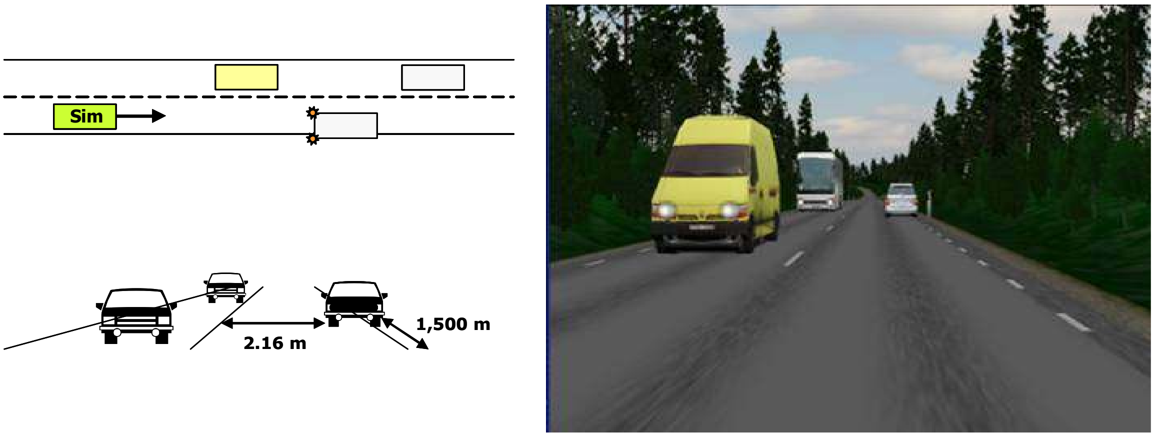


Figure 57. Pictorial view (left) and simulation screen-shot (right) of Event B from a bird's eye of view (top left) and from the driver's perspective (bottom left and right part of the figure). Drivers had to let the oncoming traffic pass before they could overtake the parking vehicle.

Manoeuvres initiated by drivers in response to Event B

Except one driver in the HC group, all drivers reacted to Event B by releasing the accelerator pedal and applying the brake and - after having let the oncoming vehicles pass - by finally accelerating and overtaking the parking vehicle. The one HC driver released the accelerator pedal and then initiated the overtaking manoeuvre without applying the brake before. One LDW driver did not drive Event B because of technical problems.

Timeliness of drivers' reaction to Event B

About one third of the drivers in each group decelerated and accelerated several times when they approached Event B. Because Event B occurred at the beginning of Lap 2 (which was the first lap driven in Session 2 for one third of the drivers), at least for some drivers this appeared to be the effect of them changing up gears in order to reach the desired speed. Because it was hard to determine whether drivers' first release of the accelerator pedal after Event B became visible was already a reaction to it, the starting point of drivers' release of the accelerator pedal was determined based on drivers' last deceleration period before they applied the brakes, although this was probably not the point when they had actually *perceived* the situation (which is a problem however with all behavioural indicators).

Drivers started to release the accelerator pedal about 187 m (*SD* = 70.40) before reaching the parking vehicle. The TTC with the second oncoming vehicle (which drivers first had let to pass) at the point where drivers started to release the accelerator pedal served as an indicator for the timeliness of drivers' reaction to Event B. The results are depicted in Figure 58.

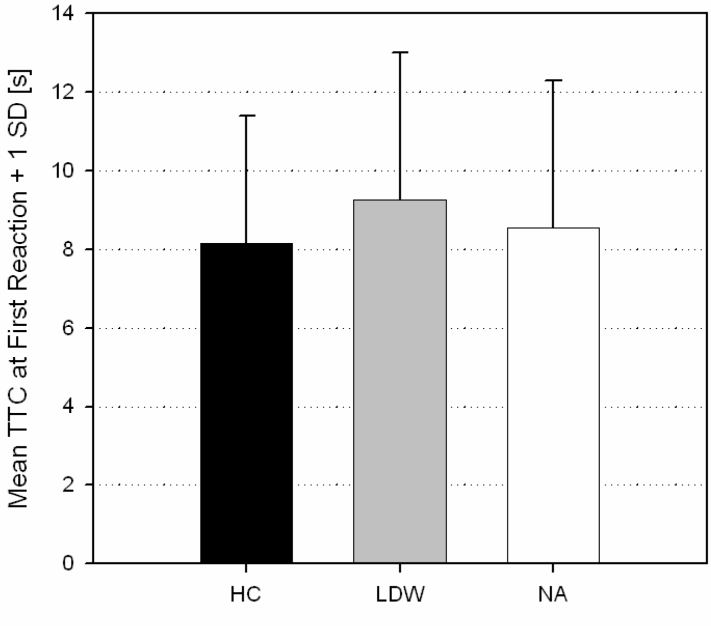


Figure 58. Mean Time to Collision (TTC) with the second oncoming vehicle [s] at the initial point of drivers' reaction to Event B across the three levels of lane keeping assistance.

As Figure 58 shows, LDW drivers started to release the accelerator pedal earlier (*M* = 9.25 s, *SD* = 3.74) than NA drivers (*M* = 8.55 s, *SD* = 3.75) and HC drivers (*M* = 8.16 s, *SD* = 3.22). As can be seen in Figure 58, there was however a large variation within groups. A one-way ANOVA yielded no significant effect of the level of lane keeping assistance on the TTC at drivers' initial reaction, *F*(2, 41) = .35, *p* = .71, η^2_G = .02.

Abruptness of initiated driving manoeuvre

Two measures were used as indicators for the abruptness of the manoeuvre initiated by drivers: (a) the mean time between drivers' release of the accelerator pedal and their application of the brake, and (b) drivers' maximum braking force.

Time between drivers' release of the accelerator pedal and their application of the brake

Two LDW drivers braked two times and accelerated in between when they approached Event B. Because the time between releasing the accelerator pedal and start braking could not be meaningfully assessed for those two drivers, they were excluded from the analysis. Furthermore, two NA drivers and one LDW driver were identified as outliers (with values laying more than 2.5 *SD* above the group means) and were therefore also excluded from the analysis.

LDW drivers took most time to apply the brake after having started to release the accelerator pedal ($M = 2.63$ s, $SD = 1.44$), followed by HC drivers ($M = 2.44$ s, $SD = 1.82$) and NA drivers ($M = 2.17$ s, $SD = 1.67$). Because the data was not normally distributed it was analysed with a non-parametric analysis of variance (Kruskal-Wallis) which yielded no significant differences between the three groups of drivers, $\chi^2(2, N = 38) = 1.71$, $p = .42$.

Maximum braking force

LDW driver were found to brake harder ($M = 131.62$ N, $SD = 48.72$) than HC drivers ($M = 128.71$ N, $SD = 52.64$); whereas NA drivers ($M = 113.55$ N, $SD = 38.59$) braked least hard when approaching Event B. However, these differences were not very pronounced and not significant as revealed by a one-way ANOVA, $F(2, 39) = .63$, $p = .54$, $\eta^2_G = .03$[27].

Influence of drivers' reliance on the lane keeping assistance systems on the timeliness and abruptness of their reaction to Event B

Correlation analyses were performed in order to reveal whether drivers' reliance on the lane keeping assistance systems (measured by their preparedness to allocate their visual attention away from the road scene during performance of the visually demanding secondary task) was related to the timeliness and the abruptness of their reaction to Event B used as indicators for drivers' SA. For this purpose, the TTC at the initial point of drivers' reaction to Event B and the time between drivers' release of the accelerator pedal and start braking were correlated with the total time that drivers had glanced to the Arrows Task display before they started to react to it[28] and with the mean duration of drivers' single display glances during the Arrows Task where they encountered Event B.

---

[27] One HC drivers was excluded from the analysis because of an extreme value that lay more than 2.5 *SD* above the group mean.

[28] The total glance time to the Arrows Task display was calculated only for drivers for whom the eye-tracking data was available for the whole duration of the Arrows Task.

A strong negative correlation was found between the TTC at the initial point of drivers' reaction to Event B and drivers' total glance time to the Arrows Task display before they started to react to Event B, $r(35) = -.77$, $p < .001$. Thus, the longer drivers had (in sum) glanced to the Arrows Task display when they approached Event B, the later did they react to it. The TTC at drivers' initial reaction was also negatively - though considerably weaker - correlated with the mean duration of drivers' single glances to the Arrows Task display during the Arrows Task where drivers encountered Event B, $r(36) = -.24$, $p = .15$.

Drivers' total glance time to the Arrows Task display before they started to react to Event B was also negatively correlated with the time between their release of the accelerator pedal and start braking, $r(32) = -.49$, $p = .004$. Thus, the longer drivers had (in sum) glanced to the Arrows Task display when they approached Event B, the more rapidly did they start to brake after they had begun to release the accelerator pedal. There was however no correlation between drivers' mean single display glance duration during the Arrows Task where drivers encountered Event B and the time between drivers' release of the accelerator pedal and start braking, $r(33) = -.03$, $p = .86$.

General characteristics of drivers' reaction to Event B

In order to reveal whether drivers in the three groups differed systematically in their reaction to Event B (which can be interpreted as an effect of the level of lane keeping assistance), some more general measures describing the characteristics of the manoeuvres initiated by drivers were analysed. For Event B, four measures were of interest: (a) the frequency of drivers' use of the turn indicator, (b) drivers' minimum speed (c) drivers' lateral position when they were at one level with the parking vehicle and (d) the starting point of drivers' initiation of the overtaking manoeuvre. Those measures did *not* serve as indicators for drivers' SA.

Drivers' use of the turn indicator

All LDW drivers used the turn indicator during their initiation of the overtaking manoeuvre, whereas three HC drivers and five NA drivers did not use the turn indicator. This difference was not significant, $p = .09$ (using Fisher's exact test), $\Phi = .35$.

Minimum speed

HC drivers had the lowest minimum speed ($M = 22.54$ km/h, $SD = 10.52$) compared to LDW drivers ($M = 27.45$ km/h, $SD = 8.60$) and NA drivers ($M = 27.57$ km/h, $SD = 6.92$) before they initiated the overtaking manoeuvre. This difference was however not significant as revealed by a one-way ANOVA, $F(2, 40) = 1.52$, $p = .23$, $\eta^2_G = .07$.

Drivers' lateral distance to the parking vehicle

Drivers' lateral distance to the parking vehicle was measured as the distance between both vehicles at the point where both vehicles were on the same level. As shown in Figure 59, HC drivers had the largest lateral distance to the parking vehicle when they were on a level with it ($M = 1.71$ m, $SD = 0.64$); however, there was also a large variance within this group of drivers ranging from a minimum distance of 0.47 m to a maximum distance of 2.52 m. LDW drivers ($M = 1.55$ m, $SD = 0.52$) did not differ much from NA drivers (*M* = 1.48 m, *SD* = 0.25) in their lateral distance to the parking vehicle. The differences between the three groups were not significant as revealed by a one-way ANOVA, *W*(2, 22.85) = .89, *p* = .42 (using Welch's statistic in order to correct for unequal variances among groups). The size of the effect was computed from the sum of squares of the ANOVA and yielded η^2_G = .04, which is a small effect.

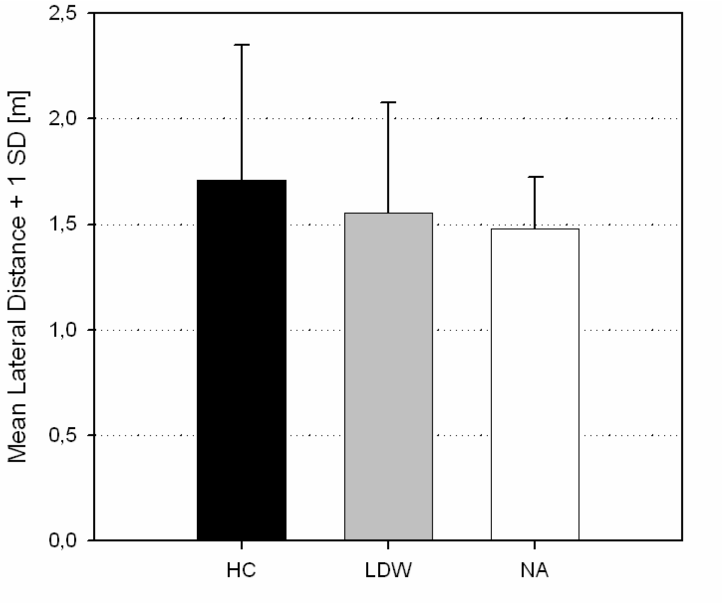


Figure 59. Drivers' mean lateral distance to the parking vehicle [m] in Event B when drivers were on a level with it as a function of the level of lane keeping assistance.

Starting point of drivers' initiation of the overtaking manoeuvre

In order to find out whether there was an effect of the level of lane keeping assistance on the relative point in time at which drivers initiated the overtaking manoeuvre, drivers in the three groups were assigned to one of two categories depending on whether they initiated the overtaking manoeuvre *before* having let passed the second oncoming vehicle, or whether they initiated the overtaking manoeuvre *after* having let passed the second oncoming vehicle. An initiation of the overtaking manoeuvre after having let passed the second oncoming vehicle was assumed to indicate a more cautious behaviour of drivers. As Figure 60 shows, more HC drivers than LDW drivers first let the oncoming traffic pass before they initiated the overtaking manoeuvre. All drivers in the control group (NA drivers) initiated the overtaking manoeuvre before they had completely let passed the oncoming traffic. The differences between the three levels of lane keeping assistance were however not significant, *p* = .10 (using Fisher's exact test), Φ = .32.

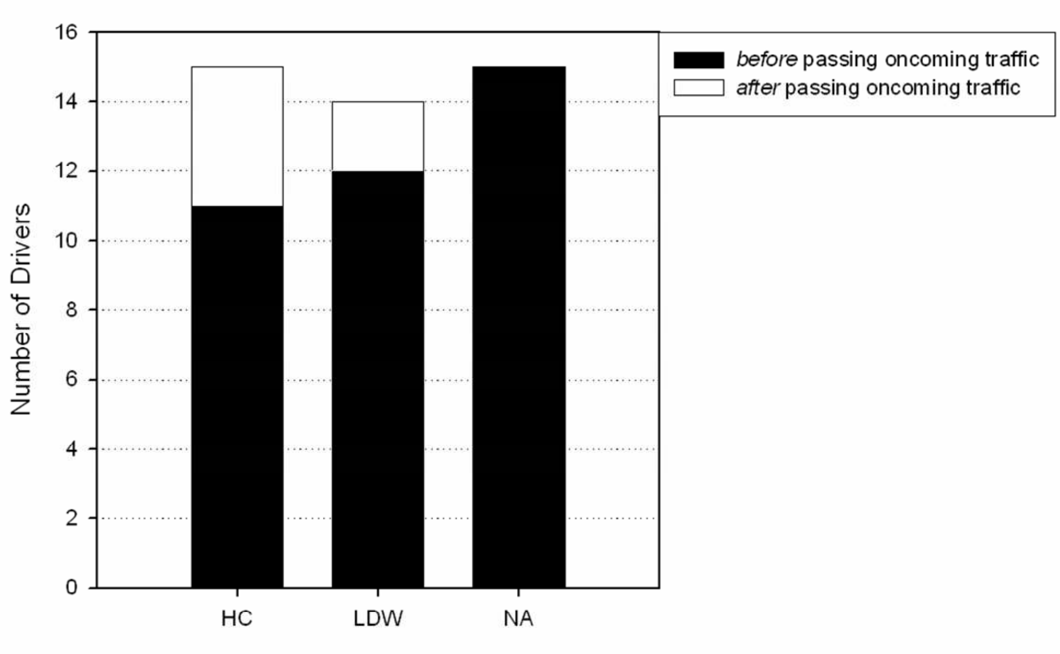


Figure 60. Number of drivers in each group who initiated the overtaking manoeuvre before vs. after having let passed the oncoming traffic.

4.6.3 Drivers' reactions to unforeseeable situations

4.6.3.1 Lateral unforeseeable – Event C

Event C was realised by three oncoming vehicles that approached drivers in a left curve whereby the vehicle in the middle position had its left turn indicator activated and partly drove on the drivers' lane as though it intended to overtake the vehicle in front of it (see Figure 61). The oncoming vehicles drove with the same speed as the drivers unless drivers' speed fell below 37 km/h at a distance headway of less than 50 m (in this case the speed of the oncoming vehicles was set to 72 km/h).

Event C was categorised as unforeseeable because drivers' sight was obstructed due to the curve. Thus, the oncoming vehicles were first visible for drivers approximately at a distance of 250 m. Drivers were not able to anticipate the Event and had a rather small time window to process the visual cues and to plan their reaction.

Drivers were required to react primarily on the lateral control level by drawing aside towards the right lane boundary in order to avoid a collision with the oncoming vehicle. Thus, Event C simulated a functional limit of the HC and LDW system.

It was hypothesised that with increasing automation of the lateral control subtask drivers allocate fewer attentional resources to the processing of visual cues relevant for lateral control of the vehicle, leading to failures in bottom-up control of attention. Those failures in bottom-up attentional control were assumed to become apparent in drivers' delayed and more abrupt reaction to this situation. Because Event C did not leave drivers much time to compensate for the hypothesised automation-induced impairments in SA, it was predicted that there would be differences in the timeliness and in the abruptness of drivers' reaction to Event C as a function of the level of lane keeping assistance. Thus, it was predicted that HC drivers show the latest and most abrupt reactions to Event C, whereas NA drivers were assumed to show the earliest and least abrupt reactions to Event C.

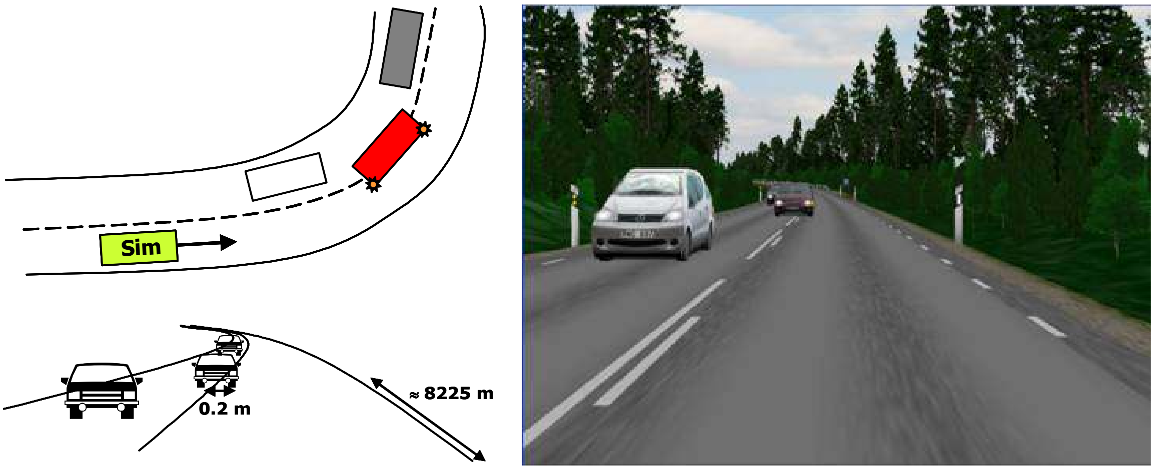


Figure 61. Pictorial view (left) and simulation screen-shot (right) of Event C from a bird's eye of view (top left) and from the driver's perspective (bottom left and right part of the figure). The oncoming vehicle in the middle position partly occupied the drivers' lane and forced drivers to draw aside in order to avoid a potential collision.

Manoeuvres initiated by drivers in response to Event C

All drivers drew aside towards the right lane boundary in reaction to Event C[29]. All drivers reacted additionally on the longitudinal control level by a reduction of speed, except for one NA driver who instead reacted by accelerating when she or he started to draw aside. Most drivers in each group released the accelerator pedal and braked, whereas only two drivers in each group only released the accelerator pedal without braking, as can be seen in Figure 62.

---

[29] One LDW driver did not drive Event C because of technical problems.

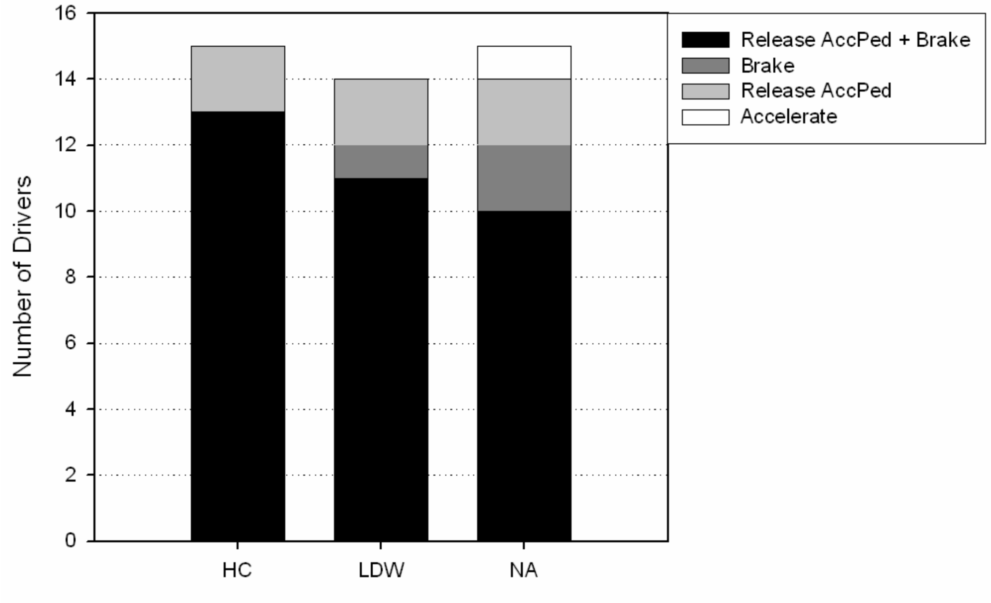


Figure 62. Type of driver reactions on the longitudinal control level in response to Event C across the three levels of lane keeping assistance (AccPed = Accelerator Pedal).

Timeliness of drivers' reaction to Event C

As shown in Figure 63, most drivers in each group first reacted to Event C by releasing the accelerator pedal, whereas a smaller amount of drivers first started to draw aside and then afterwards reacted on the longitudinal control level.

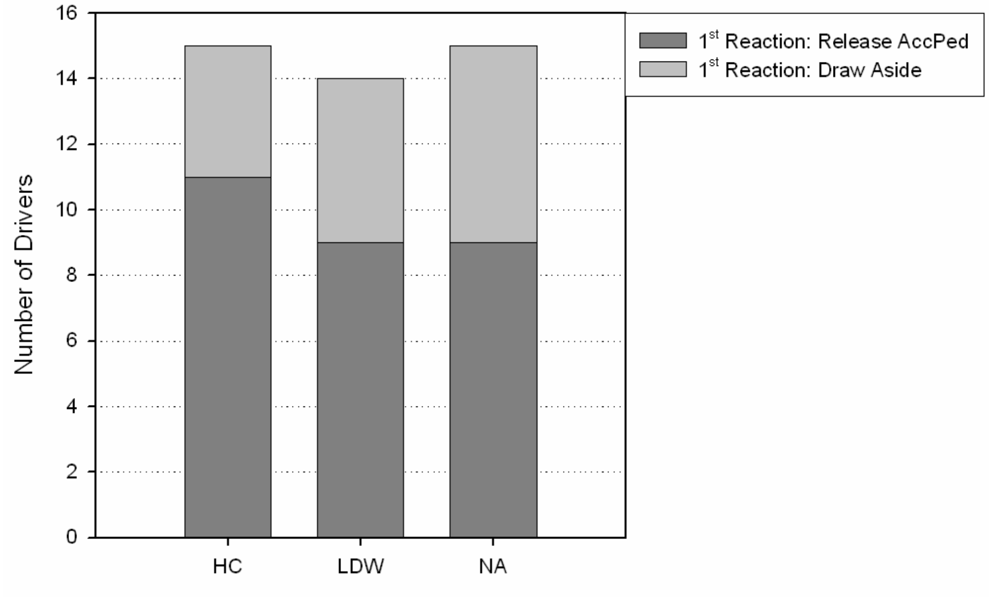


Figure 63. Drivers' first reaction to Event C across the three levels of lane keeping assistance (AccPed = Accelerator Pedal).

In order to reveal whether there was any effect of the level of lane keeping assistance on the timeliness of drivers' reaction to Event C, the three groups of drivers were compared with respect to the mean TTC with the second oncoming vehicle at the point where they initially reacted to Event C (either by releasing the accelerator pedal or by drawing aside).

Exploratory data analysis revealed that the two subgroups of HC drivers (that were found to differ significantly in their reliance on the HC system as measured by their mean single display glance durations) differed considerably in the timeliness of their reaction to Event C as measured by the TTC at the starting point of their reaction. Drivers in the 'HC short' subgroup who were found to not rely at all on the HC system reacted about 1 s earlier to Event C (*M* = 5.80 s, *SD* = 0.79) than drivers in the 'HC long' subgroup who were found to rely to a large extent on the HC system (*M* = 4.95 s, *SD* = 1.17). A t-test for independent groups yielded *t*(13) = -1.68, *p* = .12 (two-tailed), *r* = -.42.

In the subsequent analysis of the initial point of drivers' reaction to Event C the two subgroups of HC drivers were therefore treated separately. Figure 64 shows the mean TTC at drivers' initial reaction to Event C for all groups of drivers.

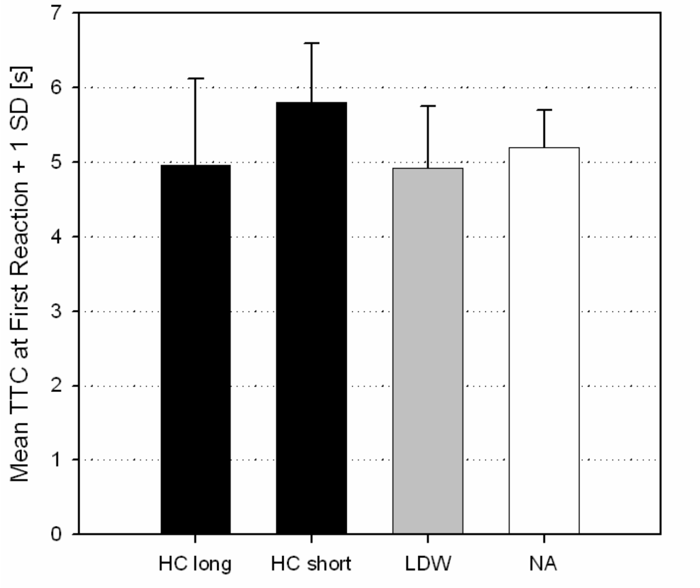


Figure 64. Mean Time to Collision (TTC) to the second oncoming vehicle [s] at the initial point of drivers' reaction to Event C for the two subgroups of HC drivers (that were found to differ considerably in their reliance on the HC system) and for LDW and NA drivers.

As can be seen in Figure 64, only drivers in the 'HC long' subgroup showed in accordance with the hypothesis a rather late reaction to Event C, whereas drivers in the 'HC short' subgroup, and not NA drivers, showed the earliest reaction to Event C. LDW drivers (*M* = 4.91 s, *SD* = 0.83) reacted contrary to the hypothesis even slightly later to Event C than drivers in the 'HC long' subgroup, whereas NA drivers (*M* = 5.19 s, *SD* = 0.50) reacted later than drivers in the 'HC long' subgroup, but earlier than LDW drivers and drivers in the 'HC long' subgroup to Event C.

A one-way ANOVA comparing the four groups of drivers yielded a marginally significant result, $F(3, 39) = 2.83$, $p = .051$, with $\eta^2_G = .17$ (corresponding a medium effect)[30]. Post-hoc tests (Tukey-HSD) confirmed that only the difference in TTC between drivers in the 'HC short' subgroup and LDW drivers was statistically significant ($p = .04$).

Abruptness of initiated driving manoeuvre

Two measures were used as indicators for the abruptness of the manoeuvre initiated by drivers: (a) the mean time between drivers' release of the accelerator pedal and their application of the brake, and (b) drivers' maximum braking force.

Time between drivers' release of the accelerator pedal and their application of the brake

Exploratory data analysis revealed that the two subgroups of HC drivers did not differ in the time between their release of the accelerator pedal and start braking. For this reason, the two HC subgroups were not treated separately in the analysis of this measure.

Figure 65 shows the results for the mean time between drivers' release of the accelerator pedal and their application of the brake as a function of the level of lane keeping assistance.

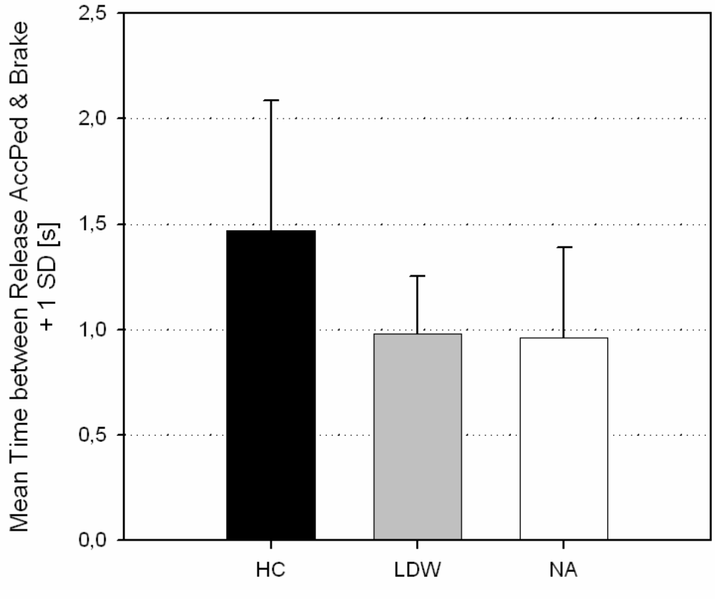


Figure 65. Mean time between drivers' release of the accelerator pedal (= AccPed) and their application of the brake [s] across the three levels of lane keeping assistance.

---

[30] One NA driver was excluded from the analysis because of an extreme value that lay more than 3 *SD* below the group mean.

As Figure 65 shows, HC drivers took considerably more time to apply the brake after having started to release the accelerator pedal (*M* = 1.47 s, *SD* = 0.62) than both LDW drivers (*M* = 0.98 s, *SD* = 0.27) and NA drivers (*M* = 0.96 s, *SD* = 0.43). A one-way ANOVA revealed that these differences between groups were marginally significant, *W*(2, 18.90) = 3.43, *p* = .054 (using Welch's statistic in order to correct for unequal variances among groups). The size of the effect was computed from the sum of squares of the ANOVA and yielded η^2_G = .22, which can be considered as large effect.

Although the two subgroups of HC drivers did not differ in the abruptness of their reaction to Event C as measured by the time between drivers' release of the accelerator pedal and start braking, it turned out that the two HC subgroups differed significantly in the mean time between their release of the accelerator pedal and the point in time at which they reached their minimum speed. Thus, drivers in the 'HC long' subgroup (*M* = 2.52 s, *SD* = 1.07) reached their minimum speed significantly faster after having started to release the accelerator pedal than drivers in the 'HC short' subgroup (*M* = 4.85 s, *SD* = 1.54), *t*(11) = -2.42, *p* = .03 (two-tailed), *r* = -.59.

The results for this additional measure of the abruptness of drivers' reaction are depicted in Figure 66, separately for the two subgroups of HC drivers. The time between drivers' release of the accelerator pedal and the point in time where they reached their minimum speed could only be calculated for drivers who released the accelerator pedal. Furthermore, the minimum speed could mostly be meaningfully assessed only for drivers who also pressed the brake pedal. This resulted in rather small case numbers as can be seen in Figure 66.

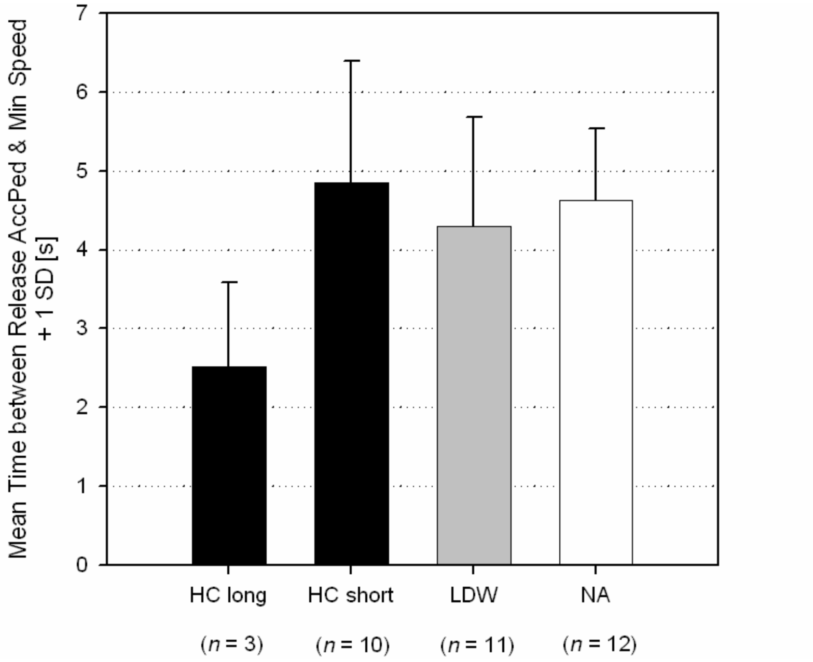


Figure 66. Mean time between drivers' release of the accelerator pedal (= AccPed) and the point in time where they reached their minimum speed [s] across the three levels of lane keeping assistance, separately for the two subgroups of HC drivers who were found to differ considerably in their reliance on the HC system.

As can be seen in Figure 66, drivers in the 'HC long' subgroup reached their minimum speed much faster after having started to release the accelerator pedal than all other groups of drivers. LDW drivers (*M* = 4.30 s, *SD* = 1.38) reached their minimum speed faster than NA drivers (*M* = 4.62 s, *SD* = 0.91). In general, the results for the mean time between drivers' release of the accelerator pedal and the point in time where they reached their minimum speed basically mirrored the pattern found for the TTC at the initial point of drivers' reaction to Event C.

A one-way ANOVA across the four groups of drivers yielded a marginally significant result, $F(3, 32) = 2.74$, $p = .06$, with $\eta^2_G = .20$ (corresponding to a rather large effect).

Maximum braking force

The two subgroups of HC drivers did not differ in their maximum braking force and were therefore not treated separately in the subsequent analysis of this measure. Although LDW drivers tended to brake harder than HC drivers and NA drivers, there were no considerable differences between LDW drivers (*M* = 118.54 N, *SD* = 72.94), HC drivers (*M* = 100.62 N, *SD* = 42.02) and NA drivers (*M* = 105.14 N, *SD* = 39.80) in terms of their maximum braking force, as also evidenced by a one-way ANOVA, $F(2, 33) = .37$, $p = .69$, $\eta^2_G = .02$.

Influence of drivers' reliance on the lane keeping assistance systems on the timeliness and abruptness of drivers' reaction to Event C

Correlation analyses were performed in order to reveal whether drivers' reliance on the lane keeping assistance systems (as measured by their preparedness to allocate their visual attention away from the road scene during performance of the visually demanding secondary task) was related to the timeliness and the abruptness of their reaction to Event C used as indicators for drivers' SA. For this purpose, the TTC at the initial point of drivers' reaction to Event C and the time between drivers' release of the accelerator pedal and start braking as well as the time between drivers' release of the accelerator pedal and the point where they reached their minimum speed were correlated with the total time that drivers had glanced to the Arrows Task display before they started to react to Event C[31] and the mean duration of drivers' single display glances measured during the Arrows Task where Event C occurred.

Interestingly, and contrary to the results of the correlation analyses for the two foreseeable situations, no correlation was found between the TTC at the initial point of drivers' reaction to Event C and drivers' total glance time to the Arrows Task display before they started to react to Event C, $r(35) = .02$, $p = .93$. Also, no correlation was found between the TTC at drivers' initial reaction and the mean duration of drivers' single glances to the Arrows Task display during the Arrows Task where drivers encountered Event C, $r(37) = .00$, $p = .98$.

---

[31] The total glance time to the Arrows Task display was calculated only for drivers for whom the eye-tracking data was available for the whole duration of the Arrows Task.

There was a small negative correlation between the time between drivers' release of the accelerator pedal and start braking and drivers' total glance time to the Arrows Task display before they started to react to Event C, *r*(26) = -.12, *p* = .56, as well as between the time between drivers' release of the accelerator pedal and start braking and the mean duration of drivers' single glances to the Arrows Task display during the Arrows Task where drivers encountered Event C, *r*(28) = -.14, *p* = .46.

The time between drivers' release of the accelerator pedal and the point in time at which they reached their minimum speed was found to be not correlated with drivers' total glance time to the Arrows Task display when they approached Event C, *r*(29) = -.06, *p* = .74. However, there was a significant negative correlation between the mean duration of drivers' single glances to the Arrows Task display during the Arrows Task where drivers encountered Event C and the time between drivers' release of the accelerator pedal and the point in time at which they reached their minimum speed, *r*(31) = -.43, *p* = .01 (two-tailed).

General characteristics of drivers' reaction to Event C

In order to reveal whether drivers in the three groups differed systematically in their reaction to Event C (which can be interpreted as an effect of the level of lane keeping assistance), some more general measures describing the characteristics of the manoeuvres initiated by drivers were analysed. For Event C, three measures were of interest: (a) the frequency of drivers' use of the turn indicator, (b) drivers' speed at the moment where they passed the oncoming vehicle partly driving on the drivers' lane and (c) drivers' lateral position when they were at one level with the oncoming vehicle. Those measures did *not* serve as indicators for drivers' SA.

Drivers' use of the turn indicator

No driver (except one HC driver) used the turn indicator during the avoidance manoeuvre in reaction to Event C.

Drivers' speed when passing the second oncoming vehicle

Drivers did not differ in their speed at the moment where they passed the second oncoming vehicle that drove partly on the drivers' lane (*M* = 49.13 km/h, *SD* = 18.49 for HC drivers; *M* = 49.27 km/h, *SD* = 21.85 for LDW drivers; and *M* = 51.69 km/h, *SD* = 17.88 for NA drivers).

Drivers' lateral distance to the second oncoming vehicle

Drivers' lateral distance to the second oncoming vehicle was measured as the distance between the two vehicles at their outer edges at the moment where they passed each other. HC drivers had on the average a slightly smaller lateral distance to the second oncoming vehicle at the moment where they passed it (*M* = 1.43 m, *SD* = 0.24) compared to NA drivers (*M* = 1.49 m, *SD* = 0.20) and LDW drivers (*M* = 1.51 m, *SD* = 0.24). A one-way ANOVA comparing the three levels of lane keeping assistance revealed that these differences were not significant, *F*(2, 41) = .56, *p* = .58, η^2_G = .03.

4.6.3.2 Longitudinal unforeseeable – Event D

Event D was realised by a vehicle that drove in front of the drivers and then suddenly set its right turn indicator, braked and turned into a lay-by at the right road side (see Figure 67). The starting point of the vehicle's braking manoeuvre was independent of drivers' distance to it, but depended on its speed and was chosen so that the vehicle would yield a definite stopping point within the lay-by with a constant deceleration of 2 m/s^2.

Because Event D required drivers to process visual cues referring to the longitudinal control level in order to react appropriately, it was predicted that the level of lane keeping assistance would not have an effect on the timeliness and the abruptness of drivers' reactions to Event D (used as indicators for drivers' SA).

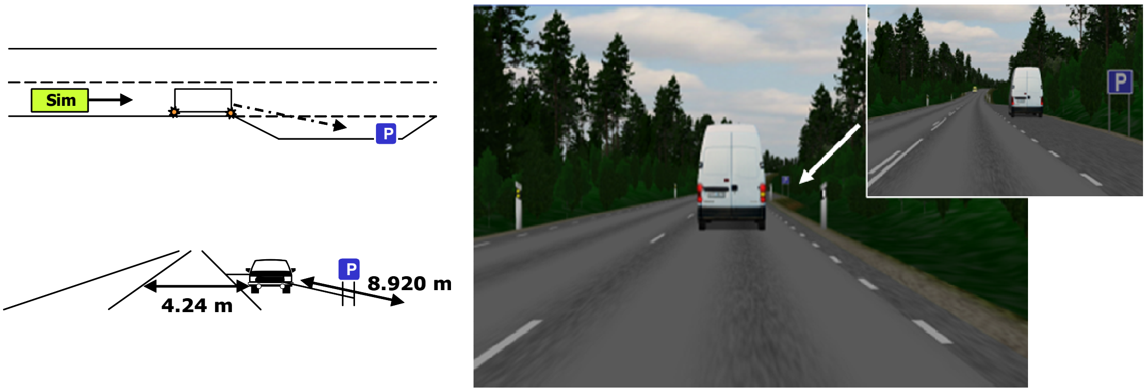


Figure 67. Pictorial view (left) and simulation screen-shot (right) of Event D from a bird's eye of view (top left) and from the driver's perspective (bottom left and right part of the figure). The vehicle in front of the driver started to brake suddenly and to turn into a lay-by at the right road side.

Manoeuvres initiated by drivers in response to Event D

Because the starting point of the vehicle's braking manoeuvre was independent of the drivers' distance to the vehicle, Event D did not work well for all drivers. Thus, five LDW drivers and three NA drivers had reduced their speed in response to the vehicle in front of them to such an extent that they were not required to react to the vehicle's braking manoeuvre because they were not driving close to it. One LDW driver did not drive this Event because of technical problems.

About half of the remaining drivers in each group reacted to Event D by first releasing the accelerator pedal and by afterwards applying the brake, whereas the other half of the drivers had already released the accelerator pedal before the vehicle started to brake and thus reacted to Event D by braking alone, as can be seen in Figure 68.

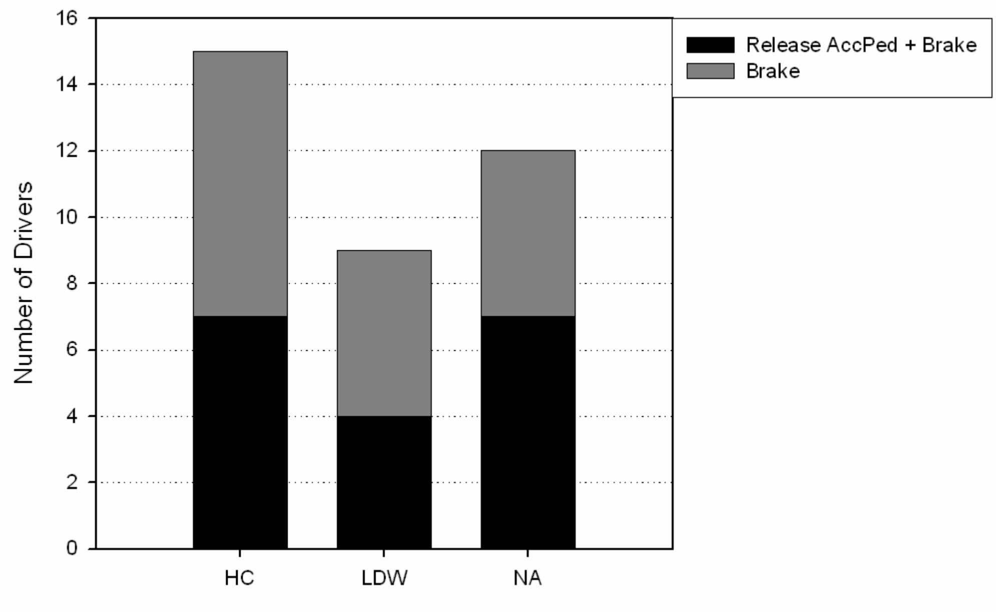


Figure 68. Drivers' reactions on the longitudinal control level in response to Event D across the three levels of lane keeping assistance. Five LDW drivers and three NA drivers had reduced their speed prior to the initiation of the vehicle's braking manoeuvre and therefore did not have to react because of a too large distance to the vehicle that initiated the braking manoeuvre (AccPed = Accelerator Pedal).

Timeliness of drivers' reaction to Event D

As a measure of the timeliness of drivers' reaction to Event D, the TTC to the vehicle in front at the point where drivers started to react to its braking manoeuvre was analysed. The results are depicted in Figure 69.

As can be seen in Figure 69, NA drivers reacted earlier (M = 18.08 s, SD = 9.92) than HC drivers (M = 11.95 s, SD = 6.76) and LDW drivers (M = 12.45 s, SD = 8.98) to Event D. A one-way ANOVA comparing the three groups of drivers yielded $F(2, 32) = 1.97$, $p = .16$, $\eta^2_G = .11$[32].

---

[32] One LDW driver was excluded from the analysis because of an extreme value that lay more than 2.5 *SD* above the group mean.

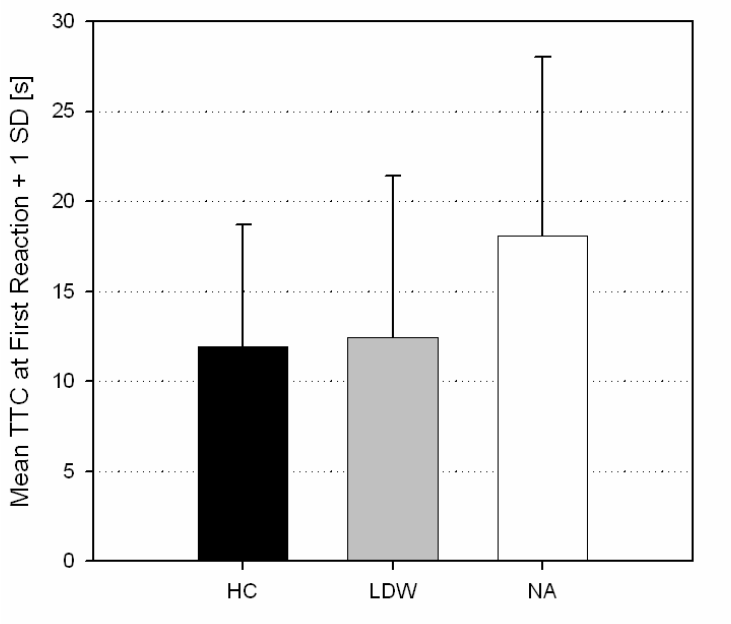


Figure 69. Mean Time to Collision (TTC) with the vehicle driving in front of drivers [s] at the initial point of drivers' reaction to its braking manoeuvre as a function of the level of lane keeping assistance.

Abruptness of initiated driving manoeuvre

Time between drivers' initial reaction and their minimum speed

Because not all drivers did react to Event D and because only about half of the remaining drivers reacted by a release of the accelerator pedal, the time between drivers' initial reaction to Event D and the point where they reached their minimum speed was taken as a measure of the abruptness of drivers' reactions (instead of the mean time between drivers' release of the accelerator pedal and their application of the brake, which would have resulted in small case numbers especially in the LDW group).

The results revealed that the mean time between drivers' first reaction to Event D and the point where they reached their minimum speed was longest for NA drivers (*M* = 2.78 s, *SD* = 1.02), followed by LDW drivers (*M* = 2.61 s, *SD* = 0.87) and HC drivers (*M* = 2.55 s, *SD* = 0.67). However, these differences between groups were small and not significant as indicated by a one-way ANOVA, *F*(2, 32) = .25, *p* = .78, η^2_G = .01[33].

Maximum braking force

Drivers did not differ in their maximum braking force (*M* = 106.87 N, *SD* = 33.09 for HC drivers; *M* = 102.64 N, *SD* = 41.02 for LDW drivers; *M* = 106.19 N, *SD* = 35.64 for NA drivers) in response to Event D, *F*(2, 33) = .04, *p* = .96, η^2_G = .00.

---

[33] One LDW driver was excluded from the analysis because of an extreme value that lay more than 2 *SD* above the group mean.

Influence of drivers' reliance on the lane keeping assistance systems on the timeliness and abruptness of drivers' reaction to Event D

Correlation analyses were performed in order to reveal whether drivers' reliance on the lane keeping assistance systems (as measured by their preparedness to allocate their visual attention away from the road scene during performance of the visually demanding secondary task) was related to the timeliness and the abruptness of their reaction to Event D used as indicators for drivers' SA. For this purpose, the TTC at the initial point of drivers' reaction to Event D and the time between drivers' first reaction and the point where they reached their minimum speed were correlated with the total time that drivers had glanced to the Arrows Task display before they started to react to Event D[34] and the mean duration of drivers' single display glances measured during the Arrows Task where Event D occurred.

No correlation was found between the TTC at the initial point of drivers' reaction to Event D and drivers' total glance time to the Arrows Task display before they started to react to Event D, $r(24) = .02$, $p = .94$[35]. There was a weak negative correlation between the TTC at drivers' initial reaction and the mean duration of drivers' single glances to the Arrows Task display during the Arrows Task where drivers encountered Event D, $r(29) = -.17$, $p = .35$.

The time between drivers' initial reaction to Event D and the point in time where they reached their minimum speed was found to be not correlated with drivers' total glance time to the Arrows Task display when they approached Event D, $r(25) = .01$, $p = .94$[36]. Also, the time between drivers' initial reaction to Event D and the point in time where they reached their minimum speed was not correlated with the mean duration of their single glances to the Arrows Task display during the Arrows Task where drivers encountered Event D, $r(30) = -.07$, $p = .70$.

4.6.4 Drivers' reactions to "anticipatory" situations

4.6.4.1 Longitudinal anticipatory – Event E

Event E was realised by a vehicle that pulled out in front of drivers from a lay-by at the right road side (see Figure 70). The vehicle started to accelerate with 2.5 m/s^2 when the drivers' Time to Collision (TTC) with it was equal to 5 s. The vehicle drove straight

---

[34] The total glance time to the Arrows Task display was calculated only for drivers for whom the eye-tracking data was available for the whole duration of the Arrows Task.

[35] Two LDW drivers were identified as outliers in the scatter plots and therefore excluded from the analysis.

[36] One LDW driver was identified as outlier in the scatter plots and therefore excluded from the analysis.

ahead in the lay-by for 2.7 s (corresponding to a distance of 10 m) before it set its left turn indicator and simultaneously started to enter the drivers' lane.

Event E was categorised as high in anticipation demands because drivers were required to anticipate the intention of the vehicle to pull out in front of them based on its acceleration within the lay-by. It was assumed that an insufficient processing of the visual cues would impair drivers' anticipation of the future development of the situation, and thus lead to delayed and more abrupt driver reactions in that situation. Because drivers had to primarily react with respect to longitudinal control to Event E (by a reduction of speed), it was predicted that there would be no differences in the timeliness and the abruptness of drivers' reaction to Event E as a function of the level of lane keeping assistance.

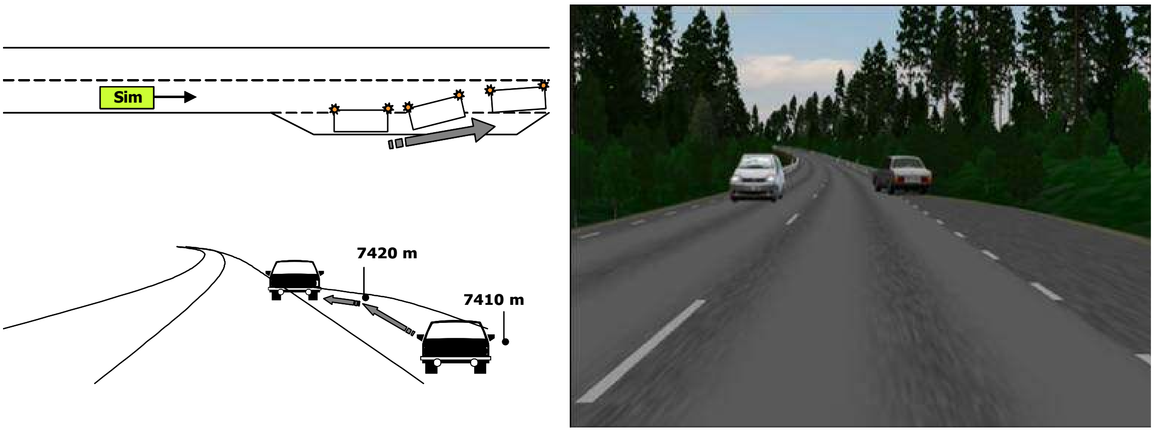


Figure 70. Pictorial view (left) and simulation screen-shot (right) of Event E from a bird's eye of view (top left) and from the driver's perspective (bottom left and right part of the figure). The vehicle starts to accelerate and to drive straight ahead in the lay-by for 2.7 s before moving towards the drivers' lane and entering the road in front of the drivers.

Manoeuvres initiated by drivers in response to Event E

An exploration of the plots depicting drivers' reactions to Event E revealed that four LDW drivers and one NA driver reacted to Event E already before the vehicle in the lay-by started to accelerate. The four LDW drivers braked a first time when their mean Time Headway (THW) to the vehicle in the lay-by was 7.84 s (*SD* = 1.21), whereas the one NA driver braked a first time at a THW of 13.77 s.

This first reaction of the five drivers that occurred prior to the Event appeared to result from drivers' anticipation that something may happen when they had noticed the vehicle standing in the lay-by. This reaction was therefore kept distinct from drivers' actual reaction to Event E that occurred in response the vehicle's acceleration and pulling-out manoeuvre. Thus, the initial point of drivers' reaction to Event E was defined as the first observable reaction that drivers showed up from 1 s before the vehicle started to accelerate within the lay-by, irrespective of whether drivers had shown a *reaction prior to Event E*.

All drivers, also the five drivers that initiated a braking manoeuvre prior to Event E, reacted to Event E either by releasing the accelerator pedal and braking, or by braking alone, as can be seen in Figure 71.

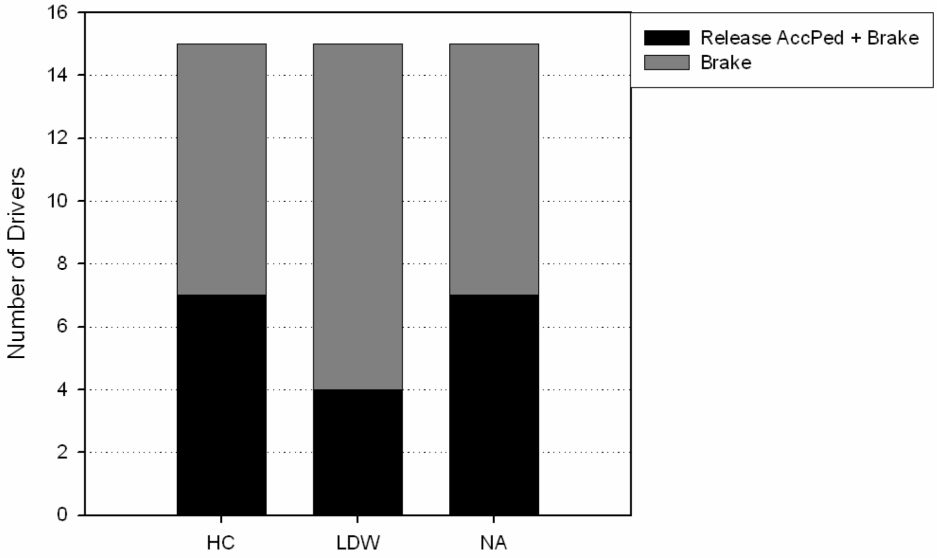


Figure 71. Drivers' reactions on the longitudinal control level in response to Event E across the three levels of lane keeping assistance (AccPed = Accelerator Pedal).

Timeliness of drivers' reaction to Event E

As a measure of the timeliness of drivers' reaction to Event E, the TTC to the pulling-out vehicle at the initial point of drivers' reaction to its acceleration and pulling-out manoeuvre was analysed. The results are depicted in Figure 72.

As shown in Figure 72, LDW drivers reacted slightly earlier ($M = 4.19$ s, $SD = 0.51$) than HC drivers ($M = 4.00$ s, $SD = 0.46$) and NA drivers ($M = 3.99$ s, $SD = 0.44$) to Event E. Those differences between groups were not significant as revealed by a one-way ANOVA, $F(2, 42) = .86$, $p = .43$, $\eta^2_G = .04$.

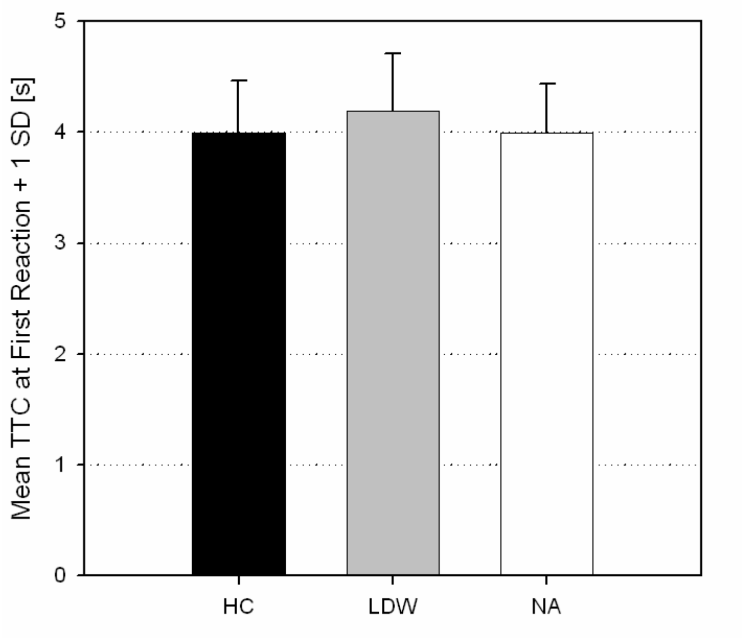


Figure 72. Mean Time to Collision (TTC) with the pulling-out vehicle [s] at the initial point of drivers' reaction to Event E across the three levels of lane keeping assistance.

Subsequent analyses revealed that the five drivers who had already shown a reaction (braking) prior to the Event reacted significantly earlier to the vehicle's acceleration within the lay-by (*M* = 4.60 s, *SD* = 0.37) than drivers who did not show a reaction prior to the Event (*M* = 3.99 s, *SD* = 0.44), *t*(43) = -2.95, *p* = .005, *r* = .41.

Abruptness of initiated driving manoeuvre

Time between drivers' initial reaction and their minimum speed

Because less than half of the drivers reacted to Event E first by a release of the accelerator pedal, the time between drivers' initial reaction to Event E (either releasing the accelerator pedal *or* braking) and the point in time where they reached their minimum speed was taken as a measure of the abruptness of their reactions.

Because the data was not normally distributed within the three groups, it was dichotomised by assigning drivers to one of two categories depending on whether the time between their initial reaction and the point in time where they reached their minimum speed was smaller vs. equal or larger than 3 s. The results are shown in Figure 73.

As illustrated in Figure 73, a majority of HC drivers and NA drivers reached their minimum speed in less than 3 s after they had started to react to Event E, whereas a majority of LDW drivers reached their minimum speed in equal or more than 3 s after they had started to react to Event E, indicating a less abrupt reaction. These differences between the three levels of lane keeping assistance were not significant, $\chi^2(2, N = 45) = 3.75$, $p = .15$, $\Phi = .29$.

The five drivers who already reacted prior to Event E all reached their minimum speed in more than 3 s after they started to react to the vehicle's acceleration within the lay-by.

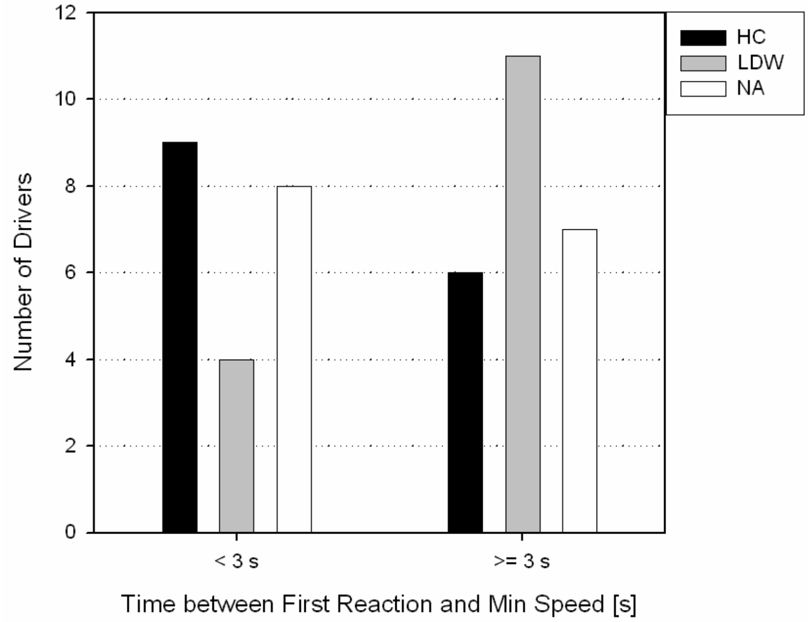


Figure 73. Number of drivers who reached their minimum speed in less than 3 s vs. in more than 3 s after their initial reaction to Event E as a function of the level of lane keeping assistance.

Maximum braking force

LDW drivers were found to brake slightly harder (*M* = 166.43 N, *SD* = 55.47) in reaction to Event E compared to HC drivers (*M* = 151.34 N, *SD* = 53.77) and NA drivers (*M* = 153.43 N, *SD* = 57.91). These differences between groups were not significant as indicated by a one-way ANOVA, $F(2, 41) = .31$, $p = .74$, $\eta^2_G = .01$[37].

Influence of drivers' reliance on the lane keeping assistance systems on the timeliness and abruptness of their reaction to Event E

Correlation analyses were performed in order to reveal whether drivers' reliance on the lane keeping assistance systems (as measured by their preparedness to allocate their visual attention away from the road scene during performance of the visually demanding secondary task) was related to the timeliness and the abruptness of their reaction to Event E used as indicators for drivers' SA. For this purpose, the TTC at the initial point of drivers' reaction to Event E and the time between drivers' first reaction and the point where they reached their minimum speed were correlated with the total time that drivers had glanced to the Arrows Task display before they started to react to

---

[37] One LDW driver was excluded from the analysis because of an extreme value that lay more than 3 *SD* above the group mean.

Event E[38] and the mean duration of drivers' single display glances during the Arrows Task where drivers encountered Event E.

No correlation was found between the TTC at the initial point of drivers' reaction to Event E and drivers' total glance time to the Arrows Task display before they started to react to the vehicle's acceleration within the lay-by, $r(36) = .07$, $p = .69$. There was a negligible positive correlation between the TTC at drivers' initial reaction and the mean duration of drivers' single glances to the Arrows Task display during the Arrows Task where drivers encountered Event E, $r(37) = .10$, $p = .52$.

Contrary to what could have been expected and also contrary to the results of the analyses of the previous Events, there was a significant positive correlation between drivers' total glance time to the Arrows Task display when they approached Event E and the time between drivers' initial reaction and the point in time where they reached their minimum speed, $r(36) = .40$, $p = .01$ (two-tailed). Thus, the longer drivers had (in sum) glanced to the Arrows Task display before they started to react to Event E, the more time passed up to the moment where they reached their minimum speed after they had started to react to Event E. There was also a significant positive correlation between the mean duration of drivers' single glances to the Arrows Task display during the Arrows Task where drivers encountered Event E and the time between drivers' initial reaction and the point in time where they reached their minimum speed, $r(37) = .32$, $p = .05$ (two-tailed).

General characteristics of drivers' reaction to Event E

In order to reveal whether drivers in the three groups differed systematically in their reaction to Event E (which can be interpreted as an effect of the level of lane keeping assistance), some more general measures describing the characteristics of the manoeuvres initiated by drivers were analysed. For Event E, two measures were of interest: (a) drivers' minimum speed and the magnitude of drivers' speed change in response to Event E and (b) drivers' minimum TTC to the pulling-out vehicle. Those measures did *not* serve as indicators for drivers' SA.

Drivers' minimum speed and the magnitude of speed change

NA drivers reduced their speed to a slightly lesser extent (M = 31.45 km/h, SD = 12.70) than HC drivers (M = 35.41 km/h, SD = 9.76) and LDW drivers (M = 35.94 km/h, SD = 12.00) in response to Event E. These differences between groups were not significant, $F(2, 41) = .67$, $p = .52$, $\eta^2_G = .03$[39]. LDW drivers reached the lowest minimum speed when they reacted to Event E (M = 41.25 km/h, SD = 9.99), followed by HC drivers (M = 43.04 km/h, SD = 11.48) and NA drivers (M = 47.05 km/h, SD = 9.85). These

---

[38] The total glance time to the Arrows Task display was calculated only for drivers for whom the eye-tracking data was available for the whole duration of the Arrows Task.

[39] One LDW driver was excluded from the analysis because of an extreme value that lay more than 2 *SD* above the group mean.

differences between groups were again quite small and not significant, $F(2, 41) = 1.18$, $p = .32$, $\eta^2_G = .05$[40].

The five drivers who already braked before the vehicle started to accelerate had a significantly lower speed ($M = 63.30$ km/h, $SD = 5.19$) when they started to react to its acceleration within the lay-by than drivers who did not show a reaction prior to Event E ($M = 79.87$ km/h, $SD = 6.91$), $t(43) = 5.16$, $p < .001$ (two-tailed), $r = .62$. In response to the vehicle's acceleration and pulling-out manoeuvre those drivers reduced their speed to a lesser extent ($M = 25.80$ km/h, $SD = 12.28$) than drivers who had not shown a reaction before the vehicle started to accelerate ($M = 35.31$ km/h, $SD = 11.04$), $t(42) = 1.79$, $p = .08$ (two-tailed), $r = .27$[41]. Nevertheless, the five drivers who showed a reaction already before the vehicle started to accelerate had a lower minimum speed ($M = 37.50$ km/h, $SD = 10.64$) than drivers who did not show a reaction prior to the Event ($M = 44.65$ km/h, $SD = 10.36$), $t(42) = 1.45$, $p = .15$ (two-tailed), $r = .22$[42].

Drivers' minimum Time to Collision (TTC) with the pulling-out vehicle

NA drivers had the smallest TTC with the vehicle during their reaction to its pulling-out manoeuvre ($M = 3.74$ s, $SD = 0.37$), followed by HC drivers ($M = 3.83$ s, $SD = 0.46$) and LDW drivers ($M = 3.89$ s, $SD = 0.46$). These differences were however small and not significant, $F(2, 41) = .41$, $p = .67$, $\eta^2_G = .02$.

The five drivers who already braked before the vehicle started to accelerate had a significantly larger minimum TTC ($M = 4.35$ s, $SD = 0.54$) with the vehicle during their reaction to its pulling-out manoeuvre than drivers who did not show a reaction prior to the Event ($M = 3.78$ s, $SD = 0.41$), $t(43) = -2.83$, $p = .007$ (two-tailed), $r = .40$.

4.6.4.2 Longitudinal anticipatory – Event F

Event F was realised by a pedestrian who crossed the road in front of the drivers and walked towards a group of people in front of a bus standing in a lay-by at the left road side (see Figure 74). The pedestrian started to walk towards the road up from a position 6.5 m right from the road when the drivers' TTC with it became equal to 6 s. The speed of the pedestrian was 1.67 m/s.

For a timely reaction to that situation drivers were required to anticipate the intention of the pedestrian to cross the road in front of them based on its movement towards the road and the group of people waiting in front of the bus. It was assumed that an insufficient processing of the visual cues would impair drivers' anticipation of the future development of the situation, and thus lead to delayed and more abrupt

---

[40] One LDW driver was excluded from the analysis because of an extreme value that lay more than 2.5 *SD* below the group mean.

[41] One LDW driver was excluded from the analysis because of an extreme value that lay more than 2 *SD* above the group mean.

[42] One LDW driver was excluded from the analysis because of an extreme value that lay more than 2.5 *SD* below the group mean.

reactions in that situation. Because drivers had to react primarily with respect to longitudinal control (by a reduction of speed), no differences were predicted in terms of the timeliness and the abruptness of drivers' reactions to Event F as a function of the level of lane keeping assistance.

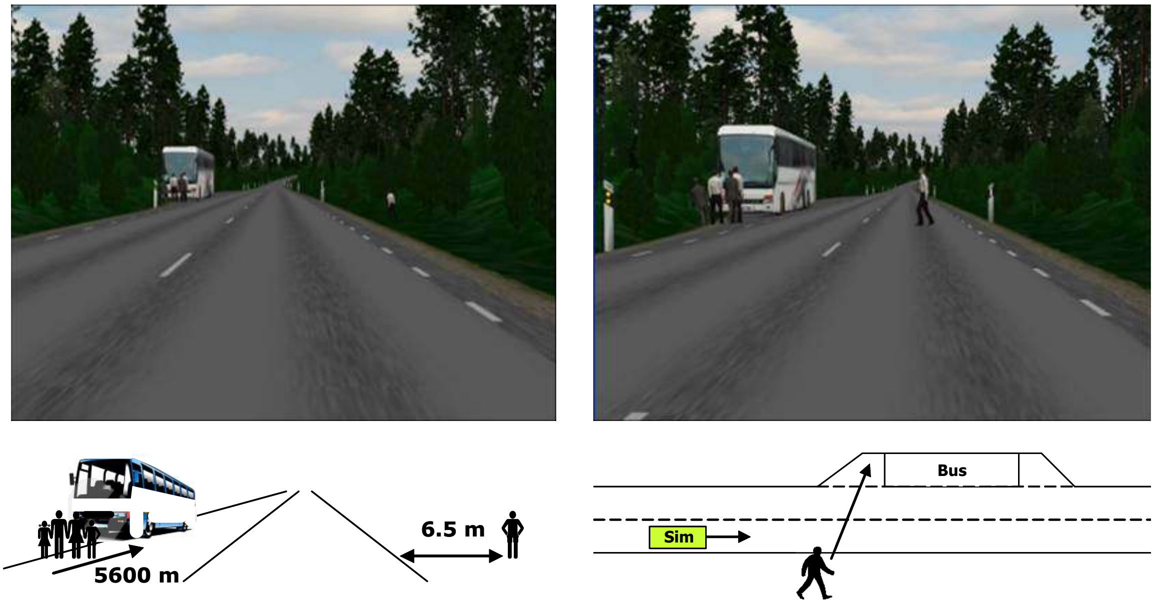


Figure 74. Top: simulation screen-shots of Event F. Top left panel: drivers' view when they approached the Event. The pedestrian started to walk at TTC = 6 s towards the road from a position of 6.5 m right from the road. Top right panel: drivers' view when they further approached the Event. The pedestrian crosses the road and walks towards the group of people positioned in front of the bus. Bottom: pictorial view of Event F from a driver's point of view (bottom left) and from a bird's eye of view (bottom right).

Manoeuvres initiated by drivers in response to Event F

Similarly to Event E, there were some drivers in each group who already reacted to Event F before the pedestrian started to walk. Thus, about one third of the drivers in each group showed a reaction on the longitudinal control level before Event F actually started, as can be seen in Figure 75. Most of these drivers released the accelerator pedal when they approached Event F, whereas one HC driver and one NA driver first released the accelerator pedal and then braked when they approached Event F. One LDW driver did not drive Event F because of technical problems. Drivers who already reacted prior to Event F released the accelerator pedal at a mean THW of 10.40 s to the pedestrian (*SD* = 1.25 s).

This first reaction that occurred prior to the Event appeared to result from drivers' anticipation that something might happen when they had noticed the bus and the group of people standing in the lay-by at the left road side. This reaction was therefore kept distinct from drivers' actual reaction to Event F that occurred in response to the movement of the pedestrian towards the road. Thus, the initial point of drivers' reaction to Event F was defined as the first observable reaction that drivers showed up from 2 s before the pedestrian started to walk towards the road, irrespective of whether drivers had shown a *reaction prior to Event F*.

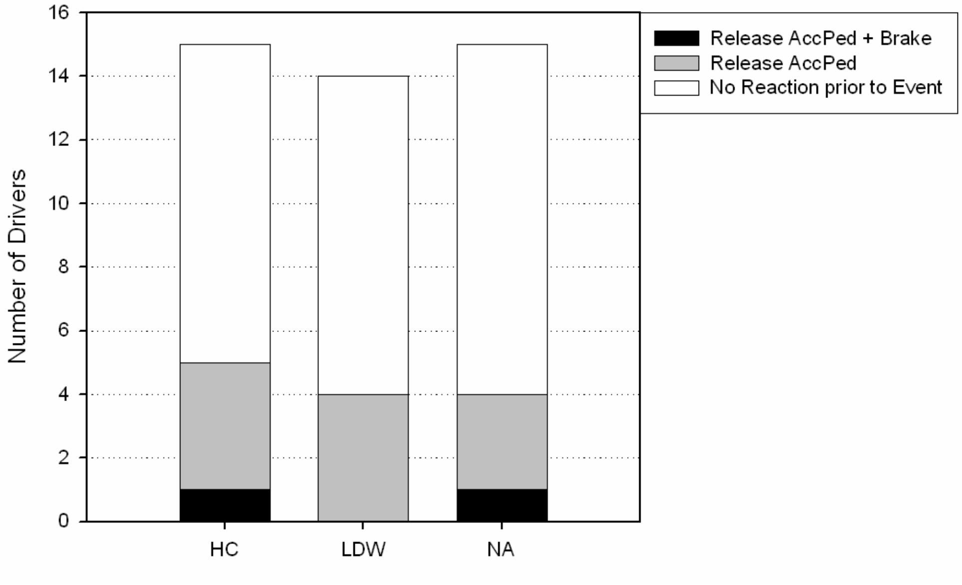


Figure 75. Number of drivers in each group who reacted prior to Event F (i.e., before the pedestrian started to walk towards the road) by either releasing the accelerator pedal (= AccPed) or by releasing the accelerator pedal and applying the brake.

All drivers reacted on the longitudinal control level to Event F (see Figure 76). A majority of the drivers in each group reacted to Event F by releasing the accelerator pedal and braking, whereas one up to two drivers in each group reacted to Event F by braking alone. One NA driver did not brake in reaction to Event F, but released the accelerator pedal and then reacted on the lateral control level by initiating an avoidance manoeuvre.

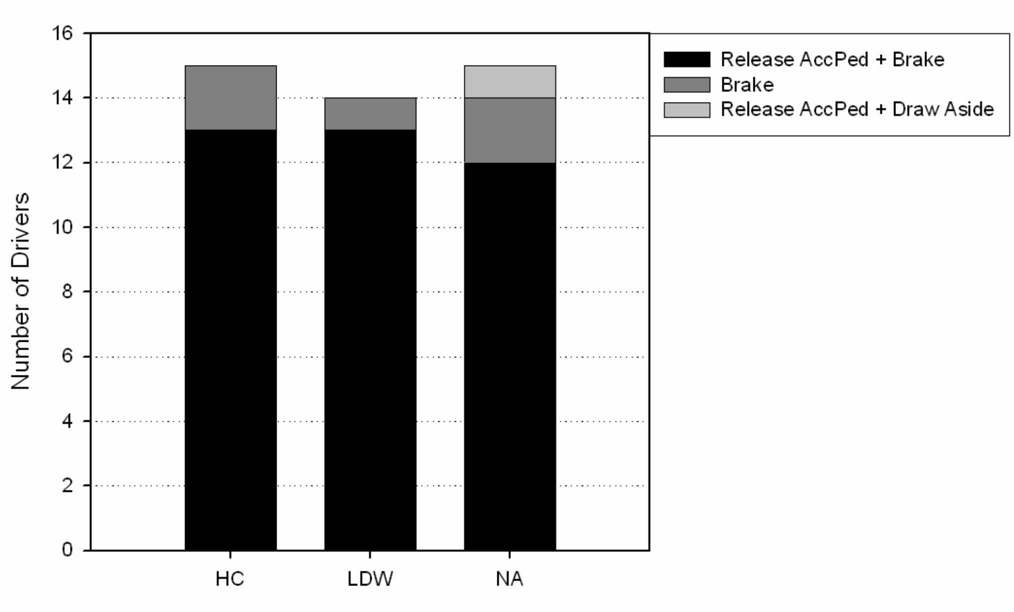


Figure 76. Type of manoeuvres initiated by drivers in each group in response to Event F up from 2 s before the pedestrian started to walk towards the road.

Timeliness of drivers' reaction to Event F

As a measure of the timeliness of drivers' reaction to Event F, the TTC to the pedestrian at the initial point of drivers' reaction (either releasing the accelerator pedal or braking) was analysed. The results are depicted in Figure 77.

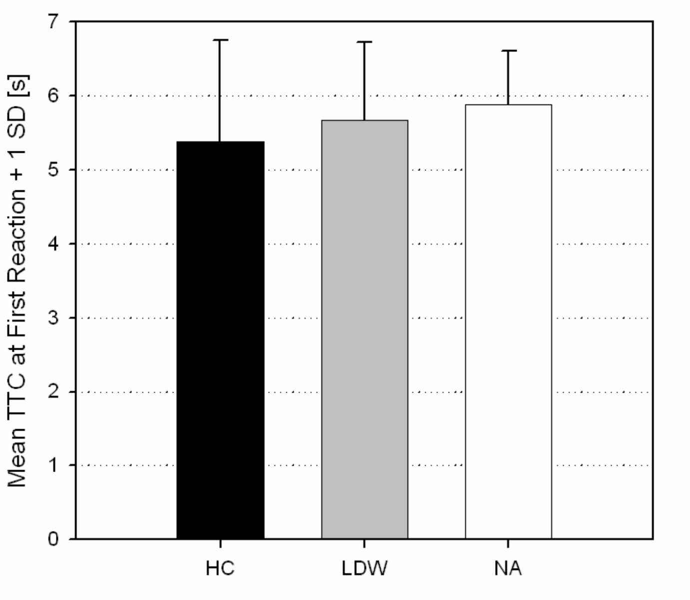


Figure 77. Mean Time to Collision (TTC) with the pedestrian crossing the road in front of the drivers [s] at the initial point of drivers' reaction to Event F as a function of the level of lane keeping assistance.

As can be seen in Figure 77, NA drivers reacted earliest to Event F (*M* = 5.87 s, *SD* = 0.73), followed by LDW drivers (*M* = 5.67 s, *SD* = 1.05) and HC drivers (*M* = 5.37 s, *SD* = 1.37). A one-way ANOVA revealed that these differences between groups were not significant, *F*(2, 40) = .77, *p* = .47, η^2_G = .04[43].

Abruptness of initiated driving manoeuvre

Time between drivers' initial reaction and their minimum speed

HC drivers reached their minimum speed fastest after they had started to react to Event F (*M* = 3.86 s, *SD* = 0.89), followed by LDW drivers (*M* = 4.15 s, *SD* = 1.52) and NA drivers (*M* = 5.16 s, *SD* = 2.13). A one-way ANOVA comparing the three groups of drivers yielded *W*(2, 24.03) = 2.33, *p* = .12 (using Welch's statistic in order to correct for unequal variances among groups). The size of the effect was computed from the sum of squares of the ANOVA and yielded η^2_G = .12 (corresponding to a medium effect).

Minimum longitudinal acceleration

Due to several extreme values in the data of drivers' maximum braking force, the results for drivers' minimum longitudinal acceleration are reported here instead. LDW drivers were found to brake hardest in response to Event F, resulting in the largest minimum longitudinal acceleration (*M* = -9.51 m/s^2, *SD* = 1.98), followed by NA drivers (*M* = -8.26 m/s^2, *SD* = 2.37) and HC drivers (*M* = -7.65 m/s^2, *SD* = 2.11). A one-way ANOVA comparing the three groups of drivers yielded *F*(2, 40) = 2.76, *p* = .08, η^2_G = .12 (corresponding to a medium effect).

Influence of drivers' reliance on the lane keeping assistance systems on the timeliness and abruptness of their reaction to Event F

Correlation analyses were performed in order to reveal whether drivers' reliance on the lane keeping assistance systems (as measured by their preparedness to allocate their visual attention away from the road scene during performance of the visually demanding secondary task) was related to the timeliness and the abruptness of their reaction to Event F used as indicators for drivers' SA. For this purpose, the THW at the initial point of drivers' reaction to Event F and the time between drivers' first reaction and the point where they reached their minimum speed were correlated with the total time that drivers had glanced to the Arrows Task display before they started to react to Event F[44] and the mean duration of drivers' single display glances measured during the Arrows Task where drivers encountered Event F.

The THW at the initial point of drivers' reaction to Event F was significantly negatively correlated with drivers' total glance time to the Arrows Task display before

---

[43] One NA driver was excluded from the analysis because of an extreme value that lay more than 2.5 *SD* above the group mean.

[44] The total glance time to the Arrows Task display was calculated only for drivers for whom the eye-tracking data was available for the whole duration of the Arrows Task.

they started to react to the movement of the pedestrian, *r*(30) = -.62, *p* < .001 (two-tailed). Thus, the longer drivers had (in sum) glanced to the Arrows Task display when they approached Event F, the later did they react to it. There was also a negative, but weaker, correlation between the THW at the initial point of drivers' reaction to Event F and the mean duration of drivers' single glances to the Arrows Task display during the Arrows Task where drivers encountered Event F, *r*(33) = -.22, *p* = .21 (two-tailed).

A significant negative correlation was found between drivers' total glance time to the Arrows Task display before they started to react to Event F and the time between drivers' initial reaction and the point in time where they reached their minimum speed, *r*(30) = -.37, *p* = .04 (two-tailed). Thus, the longer drivers had (in sum) glanced to the Arrows Task display when they approached Event F, the faster did they reach their minimum speed after they had started to react to Event F. On the contrary, there was a weak positive correlation between the mean duration of drivers' single glances to the Arrows Task display during the Arrows Task where drivers encountered Event F and the time between drivers' initial reaction and the point in time where they reached their minimum speed, *r*(33) = .12, *p* = .49 (two-tailed).

General characteristics of drivers' reaction to Event F

In order to reveal whether drivers in the three groups differed systematically in their reaction to Event F (which can be interpreted as an effect of the level of lane keeping assistance), some more general measures describing the characteristics of the manoeuvres initiated by drivers were analysed. For Event F, three measures were of interest: (a) drivers' minimum speed and the magnitude of drivers' speed change in response to Event F, (b) drivers' speed when they were at one level with the pedestrian crossing the road and (c) the position of the pedestrian when drivers were at a level with it. Those measures did *not* serve as indicators for drivers' SA.

Drivers' minimum speed and the magnitude of speed change

LDW drivers had the lowest minimum speed in reaction to Event F (*M* = 21.24 km/h, *SD* = 14.94), followed by HC drivers (*M* = 24.06 km/h, *SD* = 12.34) and NA drivers (*M* = 29.84 km/h, *SD* = 12.64). A one-way ANOVA comparing the three groups of drivers yielded *F*(2, 39) = 1.50, *p* = .23, η^2_G = .07 (corresponding to a rather small effect)[45]. In accordance with these results, LDW drivers were found to reduce their speed to the largest extent in response to Event F (*M* = 54.39 km/h, *SD* = 10.61); whereas HC drivers (*M* = 43.83 km/h, *SD* = 12.79) and NA drivers (*M* = 45.01 km/h, *SD* = 10.31) reduced their speed to a smaller extent. These differences between the three levels of lane keeping assistance were significant as revealed by a one-way ANOVA, *F*(2, 40) = 3.71, *p* = .03, η^2_G = .16 (corresponding to a medium effect). Post-hoc tests (Tukey-HSD) revealed that LDW drivers differed only significantly from HC drivers in the magnitude of their change in speed in response to Event F (*p* = .04).

---

[45] One HC driver was excluded from the analysis due to an extreme value that lay more than 2.5 *SD* above the group mean.

Drivers who already showed a reaction (braking) before the pedestrian started to walk towards the road had a significantly lower speed when they started to react to its movement later on (*M* = 68.35 km/h, *SD* = 14.94) than drivers who did not show a reaction prior to Event F (*M* = 76.05 km/h, *SD* = 7.30), $t(41) = 2.28$, $p = .03$ (two-tailed), $r = .34$. The two groups of drivers (drivers who showed vs. did not show a reaction prior to Event F) did not differ in the magnitude of their speed change in reaction to Event F, $t(41) = -.14$, $p = .89$ (two-tailed), $r = .02$.

Drivers' speed at one level with the pedestrian

HC drivers had the lowest speed when they were at a level with the pedestrian crossing the road (*M* = 42.86 km/h, *SD* = 9.39), followed by LDW drivers (*M* = 46.83 km/h, *SD* = 10.67) and NA drivers (*M* = 51.36 km/h, *SD* = 9.09). A one-way ANOVA yielded $F(2, 39) = 2.67$, $p = .08$, $\eta^2_G = .12$ (corresponding to a medium effect)[46].

Drivers who already reacted to Event F before the pedestrian started to walk had a significantly lower speed (*M* = 41.59 km/h, *SD* = 9.82) when they were at a level with it than drivers who did not show a reaction prior to Event F (*M* = 49.45 km/h, *SD* = 9.44), $t(40) = 2.46$, $p = .02$ (two-tailed), $r = .36$[47].

Position of the pedestrian when drivers were at one level with it

A majority of the drivers in each group had adjusted their speed to such an extent that the pedestrian had already crossed the road when they were at a level with it, as can be seen in Figure 78.

In all cases where the pedestrian still crossed the road when drivers were at a level with it, it had already crossed the road centre line and walked on the opposite driving lane.

---

[46] One HC driver was excluded from the analysis because of an extreme value that lay more than 2 *SD* above the group mean.

[47] One HC driver was excluded from the analysis because of an extreme value that lay more than 2 *SD* above the group mean.

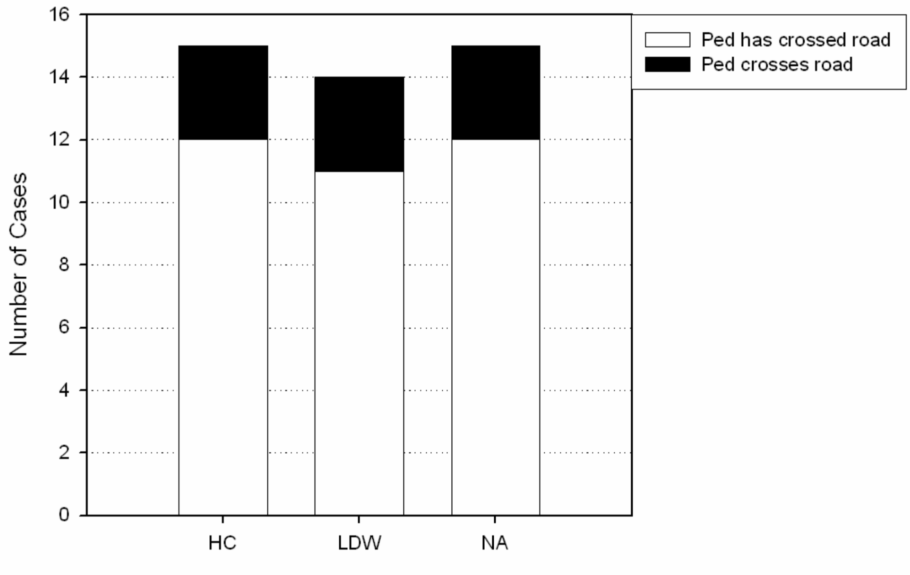


Figure 78. Number of cases in which the pedestrian (= Ped) has already crossed the road vs. still crosses the road when drivers were at a level with it across the three levels of lane keeping assistance.

4.6.4.3 Lateral anticipatory – Event G

Event G was realised by a defective truck at the left road side that partly occupied the opposing lane so that the oncoming traffic had to overtake the defective truck, and while doing this, partly occupied the drivers' lane (see Figure 79).

The defective truck was placed on the top of an ascending slope so that it was visible for drivers long (approximately 500 m) before they approached the truck. There were several oncoming vehicles that overtook the truck before drivers came close to it. Thus, if drivers perceived and interpreted the situation correctly, they were able to anticipate that they would eventually have to draw aside at the moment they passed the truck because of oncoming traffic overtaking it and partly occupying their lane. Therefore, Event G was categorised as high in anticipation demands.

There were two oncoming vehicles that overtook the truck at the moment when drivers came close to it. The two vehicles approached the truck first with the same speed as the drivers, and continued to drive with a fixed speed (36 km/h) as soon as their distance to the drivers became lower than 150 m. Thus, the actual position of the two vehicles when drivers approached the truck could vary depending on drivers' speed.

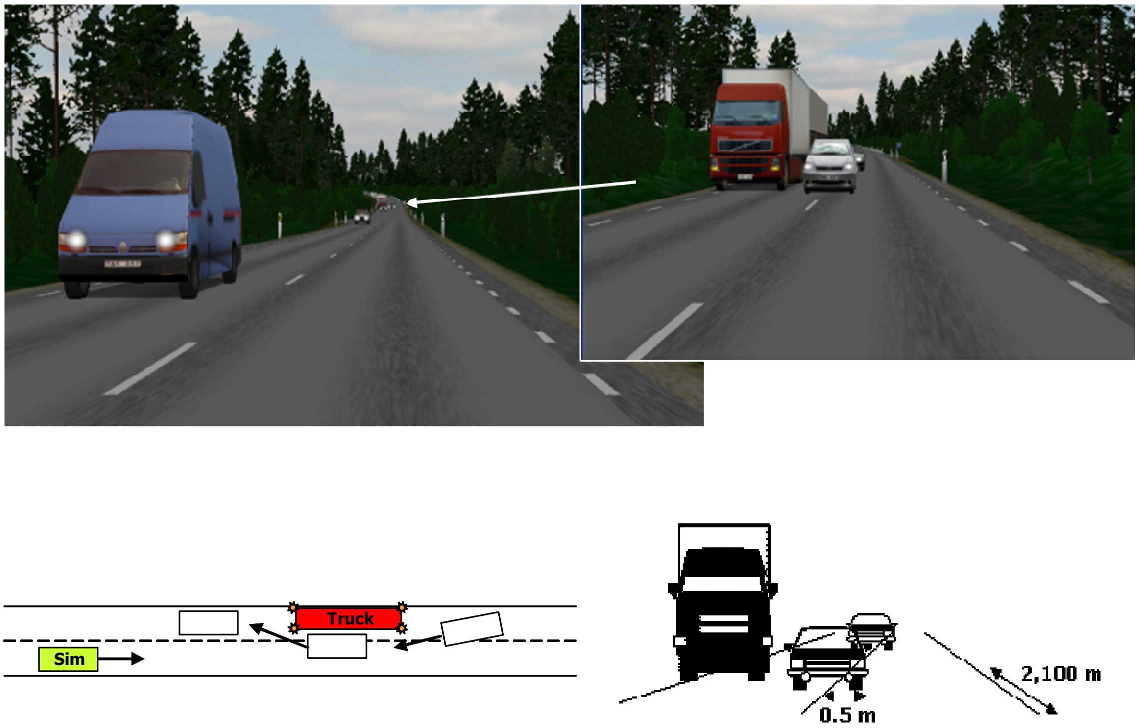


Figure 79. Top: simulation screen-shots of Event G from a driver's view about 350 m before passing the defective truck on the left roadside (top left panel) and just before approaching the truck with two oncoming overtaking vehicles partly occupying the drivers' lane (top right panel). Bottom: pictorial views of Event G from a bird's eye of view (bottom left panel) and from a driver's point of view (bottom right panel).

Drivers were required to react primarily on the lateral control level in order to avoid a collision with the oncoming vehicles. It was hypothesised that with increasing automation of the lateral control subtask drivers allocate fewer attentional resources to the processing of visual cues relevant for lateral control of the vehicle. It was assumed that insufficient processing of the visual cues would lead to an incomprehensive incorporation of that information in the situation model (lower SA), thus making anticipation difficult and leading to delayed and more abrupt driver reactions in that situation. It was therefore predicted that HC drivers would suffer to the greatest extent from automation-induced impairments in SA, and thus show the latest and most abrupt reactions in this situation, whereas NA drivers were expected to react earliest and least abruptly to Event G.

Manoeuvres initiated by drivers in response to Event G

A visual exploration of the plots depicting drivers' reactions to Event G revealed that about 80% of the drivers in each group reacted a first time to Event G about 350 m before they passed the defective truck. At this distance drivers could see some oncoming vehicles that were just overtaking the defective truck on the opposing lane (see top left panel of Figure 79). Drivers reacted to that by a reduction of speed, as can be seen in Figure 80. Half of the drivers who reacted prior to Event G released the accelerator pedal, whereas the other half of the drivers released the accelerator pedal and braked. This first reaction of drivers about 350 m before passing the truck could presumably be attributed to drivers' "back-pedalling" in order to correctly interpret the approaching situation. This first reaction of drivers about 350 m before passing the truck will therefore be subsequently referred to as *anticipatory reaction*.

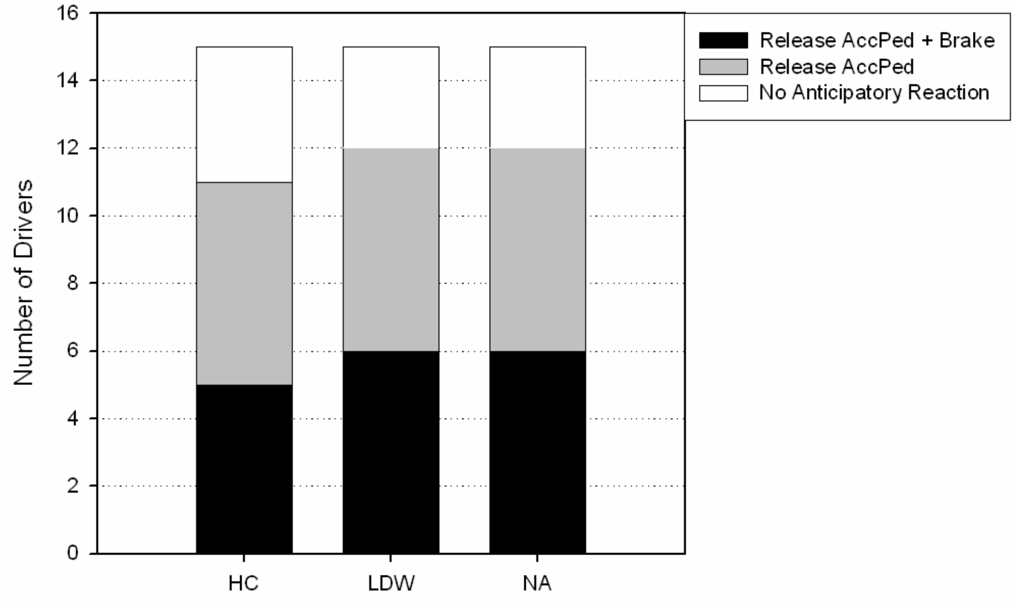


Figure 80. Number of drivers in each group who showed an anticipatory reaction (speed reduction) to Event G about 350 m before passing the defective truck (AccPed = Accelerator Pedal).

This anticipatory reaction was kept distinct from drivers' actual reaction to Event G in response to the two oncoming vehicles that drove partly on the drivers' lane when drivers approached the defective truck (see top right panel of Figure 79). Thus, the initial point of drivers' reaction to Event G was defined as the first observable reaction of drivers in response to the two oncoming vehicles partly occupying their lane just before drivers passed the defective truck.

The types of manoeuvres initiated by drivers in response to the two oncoming vehicles partly occupying their lane are depicted in Figure 81. Almost all drivers reacted both on the longitudinal control level *and* on the lateral control level to Event G. A majority of the drivers in each group reacted to Event G by releasing the accelerator pedal, braking, and drawing aside. Two LDW drivers and one NA driver just drew aside without reducing their speed, whereas one HC driver and one LDW driver reacted by a reduction of speed alone and let the oncoming traffic pass before they passed the truck.

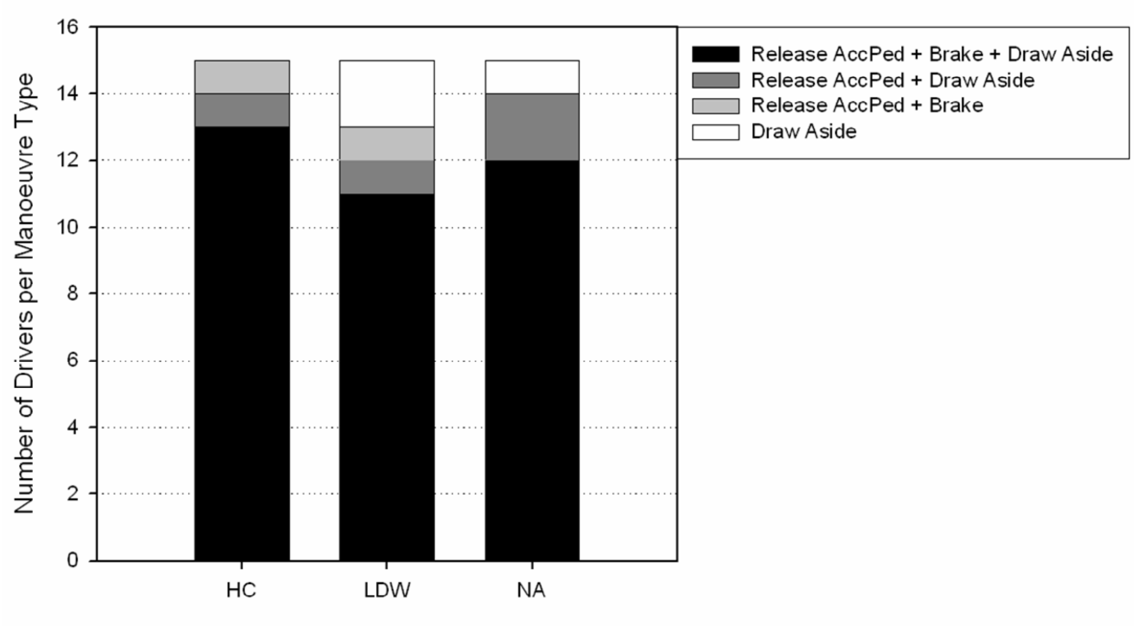


Figure 81. Manoeuvres initiated by drivers just before they passed the defective truck across the three levels of lane keeping assistance (AccPed = Accelerator Pedal).

Timeliness of drivers' reaction to Event G

As shown in Figure 82, most drivers in each group first reacted to Event G by releasing the accelerator pedal, whereas a smaller number of drivers first started to draw aside (and partly also reacted on the longitudinal control level afterwards).

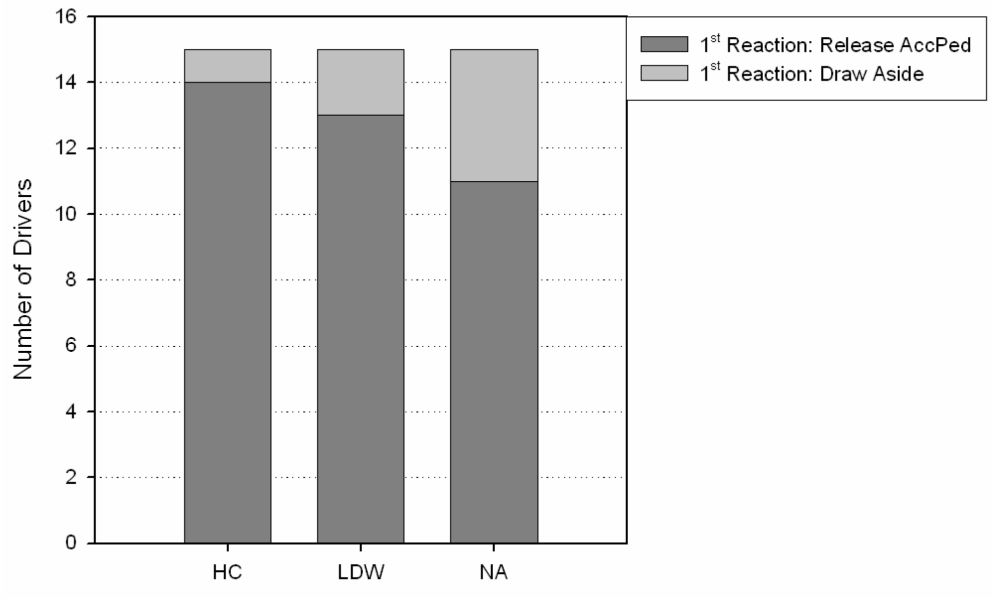


Figure 82. Drivers' first reaction to Event G across the three levels of lane keeping assistance (AccPed = Accelerator Pedal).

In order to reveal whether there was any effect of the level of lane keeping assistance on the timeliness of drivers' reaction to Event G, the three groups of drivers were compared with respect to their mean TTC with the first oncoming vehicle at the point where they initially reacted to Event G (either by releasing the accelerator pedal or by drawing aside).

Exploratory data analysis revealed that the data of the TTC at drivers' initial reaction to Event G was not normally distributed. Therefore, the data was dichotomised by assigning drivers to one of two categories depending on whether they started to react to Event G at a TTC of less than 4.5 s (corresponding to a late reaction) vs. at a TTC of 4.5 s or more (corresponding to an early reaction). The results are depicted in Figure 83.

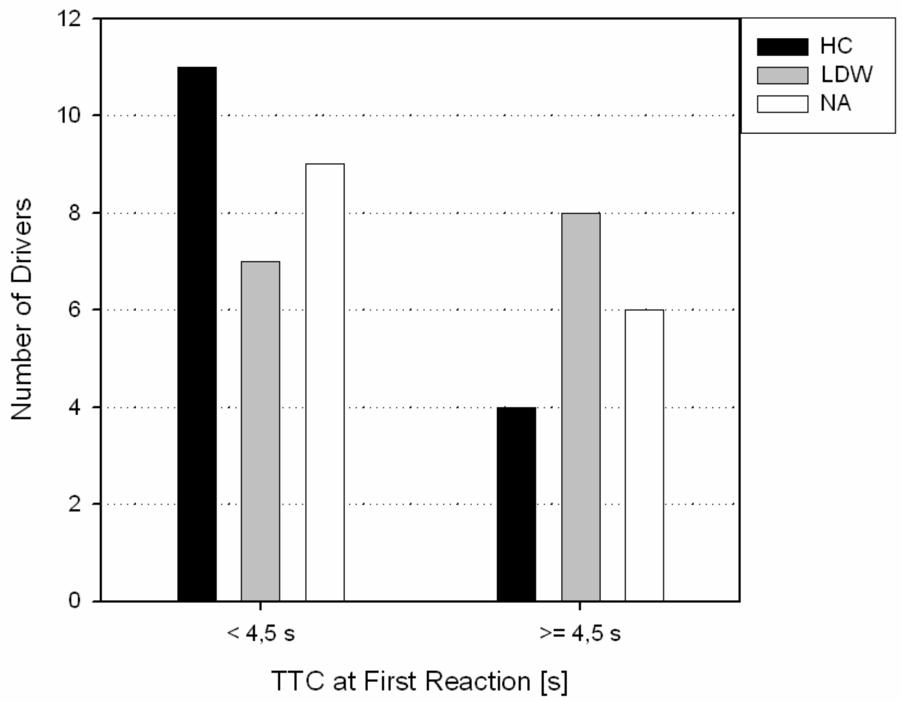


Figure 83. Number of drivers in each group who started to react to Event G at a Time to Collision (TTC) < 4.5 s vs. at a TTC >= 4.5 s with the first oncoming vehicle.

As can be seen in Figure 83, in accordance with the hypotheses, HC drivers reacted latest to Event G, evidenced by a higher proportion of drivers in this group who started to react to Event G at a TTC of less than 4.5 s. Contrary to the predictions, LDW drivers and not NA drivers reacted earliest to Event G, as evidenced by a higher proportion of drivers in the LDW group who started to react to Event G at a TTC of 4.5 s or more. However, these differences between the three levels of lane keeping assistance were rather small and turned out to be not significant, $\chi^2(2, N = 45) = 2.22$, $p = .33$, $\Phi = .22$.

Drivers who showed an anticipatory reaction about 350 m before they passed the defective truck started to react significantly earlier in terms of TTC with the first oncoming vehicle ($M = 4.77$ s, $SD = 0.94$) than drivers who did not show an anticipatory reaction prior to Event G (*M* = 4.09 s, *SD* = 0.45), *t*(31.96) = -3.15, *p* = .003 (two-tailed), *r* = .32 (*t* and *df* adjusted for unequal variances among groups).

Abruptness of initiated driving manoeuvre

Two measures were used as indicators for the abruptness of the manoeuvres initiated by drivers: (a) the mean time between drivers' release of the accelerator pedal and their application of the brake, and (b) drivers' maximum braking force.

Time between drivers' release of the accelerator pedal and their application of the brake

Because the data of the time between drivers' release of the accelerator pedal and start braking was not normally distributed, it was dichotomised by assigning drivers to one of two categories depending on whether they started to brake in less than 1.5 s vs. in 1.5 s ore more after having started to release the accelerator pedal.

The results are depicted in Figure 84.

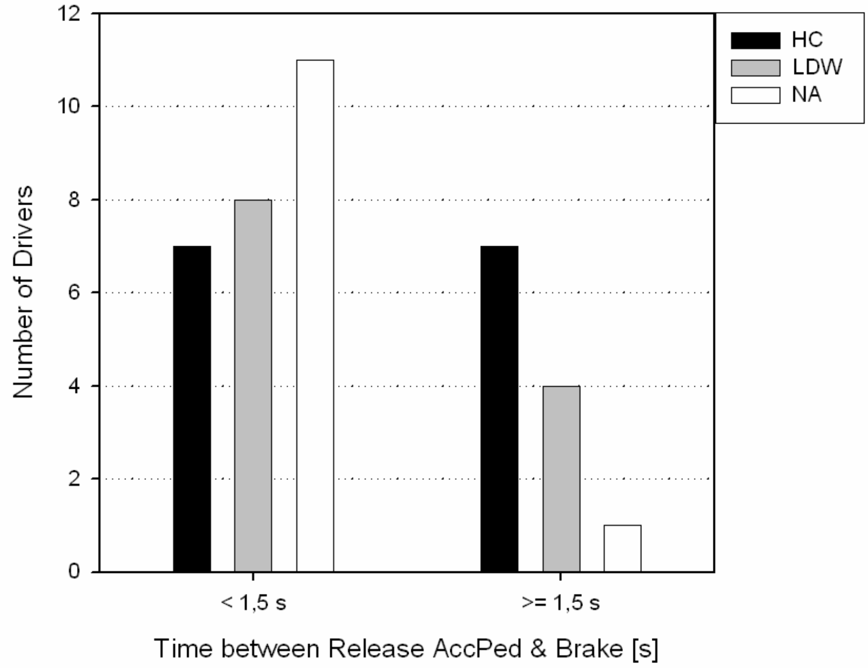


Figure 84. Number of drivers in each group who started to brake in < 1.5 s vs. in >= 1.5 s after having started to release the accelerator pedal (= AccPed).

Contrary to the predictions, NA drivers reacted most abruptly to Event G (evidenced by a higher proportion of drivers who started to apply the brake in less than 1.5 s after having started to release the accelerator pedal), whereas HC drivers reacted least abruptly to Event G (evidenced by a higher proportion of drivers who started to apply the brake in 1.5 s or more after having started to release the accelerator pedal). The differences between the three levels of lane keeping assistance were not significant, $\chi^2(2, N = 38) = 5.22$, *p* = .07 (two-tailed), Φ = .37.

Maximum braking force

NA drivers braked harder in response to Event G (*M* = 146.10 N, *SD* = 75.34), than HC drivers (*M* = 128.64 N, *SD* = 50.70) and LDW drivers (*M* = 124.94 N, *SD* = 44.44). These differences between groups were not significant as revealed by a one-way ANOVA, $F(2, 34) = .45$, $p = .64$, $\eta^2_G = .03$[48].

Drivers who showed an anticipatory reaction about 350 m before they passed the defective truck braked significantly harder (*M* = 144.07 N, *SD* = 57.61) than drivers who did not show an anticipatory reaction prior to Event G (*M* = 86.66 N, *SD* = 24.70), $t(23.17) = -4.08$, $p < .001$ (two-tailed), $r = .40$ (*t* and *df* adjusted for unequal variances among groups).

Influence of drivers' reliance on the lane keeping assistance systems on the timeliness and abruptness of drivers' reaction to Event G

Correlation analyses were performed in order to reveal whether drivers' reliance on the lane keeping assistance systems (as measured by their preparedness to allocate their visual attention away from the road scene during performance of the visually demanding secondary task) was related to the timeliness and the abruptness of their reaction to Event G used as indicators for drivers' SA. For this purpose, the TTC with the first oncoming vehicle at the initial point of drivers' reaction to Event G and the time between drivers' release of the accelerator pedal and start braking were correlated with the total time that drivers had glanced to the Arrows Task display before they started to react to Event G[49] and the mean duration of drivers' single display glances during the Arrows Task where drivers encountered Event G.

There was a weak negative correlation between the TTC at the initial point of drivers' reaction to Event G and drivers' total glance time to the Arrows Task display before they started to react to the two oncoming vehicles, $r(36) = -.15$, $p = .36$ (two-tailed)[50]. There was also a weak negative correlation between the TTC at the initial point of drivers' reaction to Event G and the mean duration of drivers' single glances to the Arrows Task display during the Arrows Task where drivers encountered Event G, $r(37) = -.11$, $p = .52$ (two-tailed)[51].

A significant negative correlation was found between drivers' total glance time to the Arrows Task display before they reacted to the two oncoming vehicles and the time between drivers' release of the accelerator pedal and start braking, $r(29) = -.44$, $p = .01$

---

[48] One LDW driver was excluded from the analysis because of an extreme value that lay more than 2 *SD* above the group mean.

[49] The total glance time to the Arrows Task display was calculated only for drivers for whom the eye-tracking data was available for the whole duration of the Arrows Task.

[50] One NA driver was identified as outlier in the scatter plots and therefore excluded from the analysis.

[51] One NA driver was identified as outlier in the scatter plots and therefore excluded from the analysis.

(two-tailed)[52]. Thus, the longer drivers had (in sum) glanced to the Arrows Task display before they passed the truck, the faster did they apply the brake after having started to release the accelerator pedal. Only a weak negative correlation was found between the mean duration of drivers' single glances to the Arrows Task display during the Arrows Task where drivers encountered Event G and the time between drivers' release of the accelerator pedal and start braking, $r(29) = -.19$, $p = .31$ (two-tailed)[53].

General characteristics of drivers' reaction to Event G

In order to reveal whether drivers in the three groups differed systematically in their reaction to Event G (which can be interpreted as an effect of the level of lane keeping assistance), some more general measures describing the characteristics of the manoeuvres initiated by drivers were analysed. For Event G, three measures were of interest: (a) the magnitude of drivers' speed change in response to Event G and drivers' speed when they were at one level with the first oncoming vehicle, (b) drivers' lateral distance to the first oncoming vehicle when they were at a level with it and (c) drivers' use of the turn indicator. Those measures did *not* serve as indicators for drivers' SA.

Drivers' speed at one level with the oncoming traffic and the magnitude of drivers' speed change

HC drivers had the lowest speed ($M = 39.83$ km/h, $SD = 20.67$) when they were at a level with the first oncoming vehicle, followed by LDW drivers ($M = 44.59$ km/h, $SD = 19.42$) and NA drivers ($M = 53.83$ km/h, $SD = 19.12$). A one-way ANOVA revealed that these differences between groups were not significant, $F(2, 42) = 1.95$, $p = .15$, $\eta^2_G = .08$ (corresponding to a small effect). HC drivers were also found to reduce their speed to the greatest extent ($M = 36.65$ km/h, $SD = 18.48$) up from their initial reaction to Event G to the point in time when they were at a level with the first oncoming vehicle, whereas LDW drivers ($M = 25.51$ km/h, $SD = 17.20$) and NA drivers ($M = 23.34$ km/h, $SD = 19.40$) were found to reduce their speed to a smaller extent in response to Event G. These differences between groups were again not significant, $F(2, 41) = 2.24$, $p = .12$, $\eta^2_G = .10$[54].

Drivers' lateral distance to the first oncoming vehicle

Drivers' lateral distance to the first oncoming vehicle was measured as the distance between the two vehicles at their outer edges at the moment where they passed each other. The results are depicted in Figure 85.

As can be seen in Figure 85, HC drivers had the smallest lateral distance ($M = 1.17$ m, $SD = 0.24$) to the first oncoming vehicle when they were at a level with it, followed

---

[52] One LDW driver and one NA driver were identified as outlier in the scatter plots and therefore excluded from the analysis.

[53] One LDW driver and one NA driver were identified as outlier in the scatter plots and therefore excluded from the analysis.

[54] One LDW driver was excluded from the analysis because of an extreme value that lay more than 2.5 *SD* above the group mean.

by LDW drivers (*M* = 1.23 m, *SD* = 0.20) and NA drivers (*M* = 1.38 m, *SD* = 0.21). These differences between the three levels of lane keeping assistance were significant as indicated by a one-way ANOVA, *F*(2, 41) = 3.62, *p* = .04, η^2_G = .15 (corresponding to a medium effect)[55]. Post-hoc tests (Tukey-HSD) revealed that only the difference between HC drivers' and NA drivers' mean lateral distance was significant (*p* = .03).

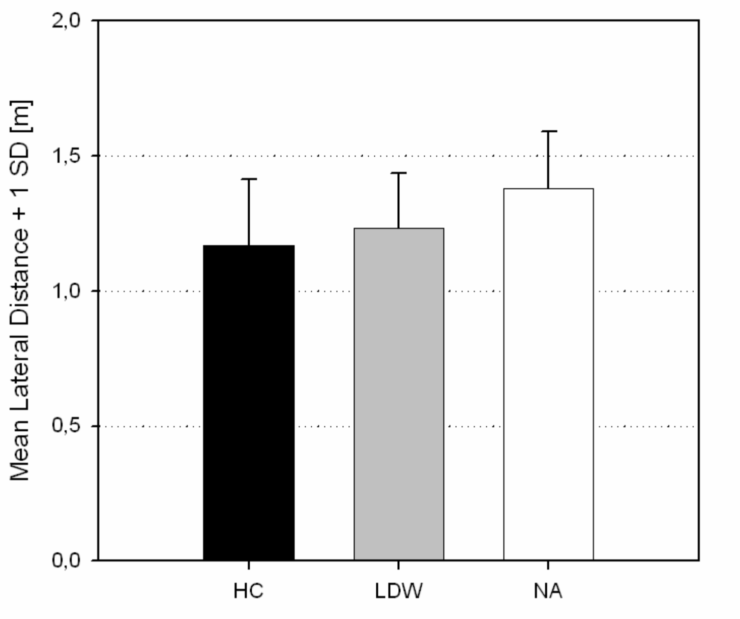


Figure 85. Drivers' mean lateral distance to the first oncoming vehicle [m] when they were at one level with it as a function of the level of lane keeping assistance.

Drivers' use of the turn indicator

Two HC drivers used the turn indicator during their avoidance manoeuvre in response to the two oncoming vehicles occupying their lane when they were just passing the defective truck. Any of the LDW drivers and NA drivers used the turn indicator.

4.6.5 Drivers' reactions to the "car-following" situation

Event H was the only situation that was categorised as "not safety-critical" in terms of demanding a reaction of drivers in order to avoid a potential collision. Event H

---

[55] One LDW driver was excluded from the analysis because of an extreme value that lay more than 2 *SD* below the group mean.

instead was designed to assess drivers' SA on the basis of their tendency to follow the erroneous behaviour of a lead vehicle which broke a speed limit (see Figure 86).

The lead vehicle drove for about 1 km in front of drivers with the same speed as the drivers. When the vehicle passed a 50 km/h speed sign, it did not reduce its speed accordingly, but continued to drive with a constant speed that it had before passing the sign. After 600 m there was another speed sign that set back the speed limit to 70 km/h.

Event H was the only situation in which drivers' SA was *not* assessed by the timeliness and the abruptness of drivers' reactions, but by drivers' tendency to show a reaction in response to the 50 km/h speed limit or not. No effect of the level of lane keeping assistance was expected in this situation.

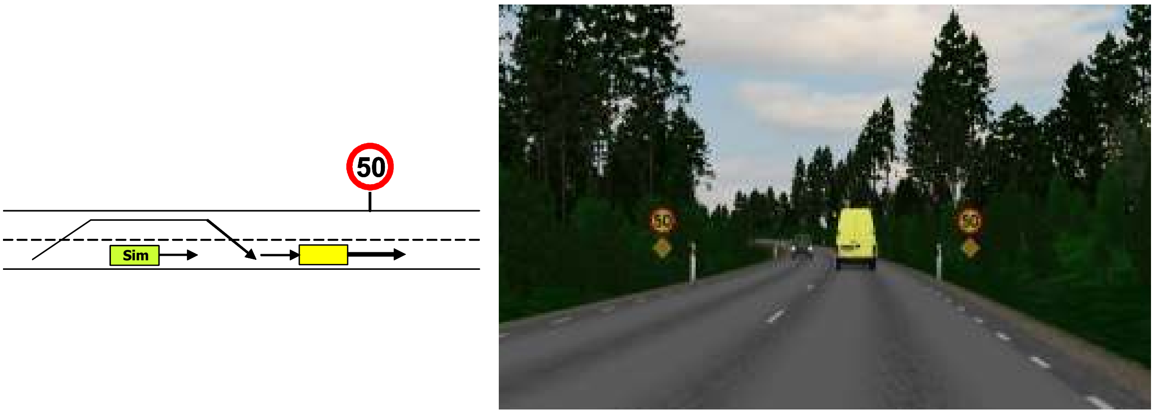


Figure 86. Pictogram from a bird's eye view (left) and simulation screen-shot (right) of Event H. The yellow vehicle in front of the drivers violates the 50 km/h speed limit and continues to drive with a higher speed.

In order to reveal whether drivers were *aware* of the changing speed limit, any reaction (either a clear reduction of speed by releasing the accelerator pedal and/or braking, or an absent acceleration before the subsequent 70 km/h speed sign) of drivers was counted regardless of how fast or accurately it was executed.

The results revealed that almost all drivers were aware of the changing speed limit, as can be seen in Figure 87. One LDW driver had reduced his or her speed to less than 50 km/h in response to lead vehicle already before passing the 50 km/h speed sign so that a reaction could not be clearly identified. One further LDW driver did not drive Event H because of technical problems. Out of the remaining drivers, only one HC driver and one NA driver did not react to the changing speed limit indicating that they were not aware of it.

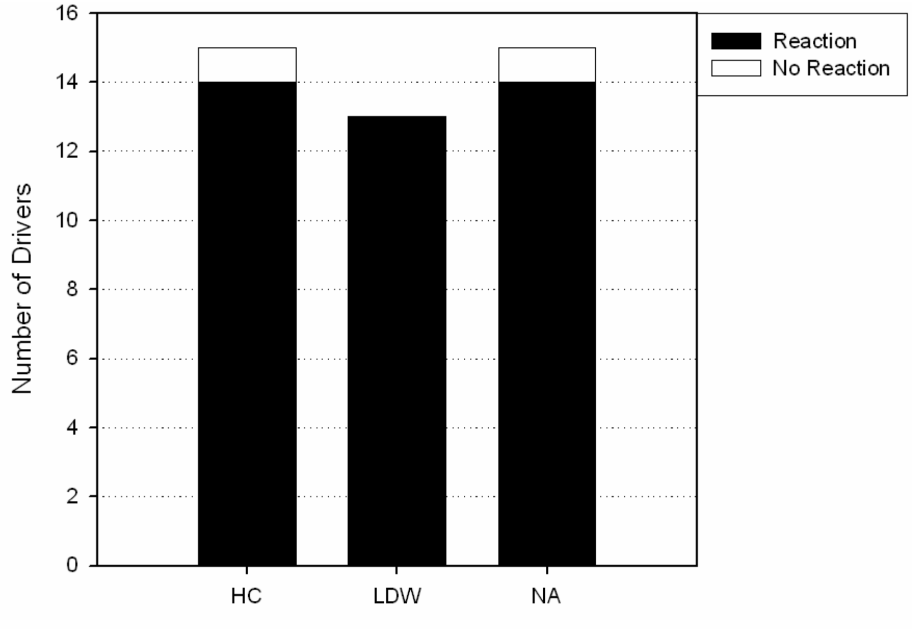


Figure 87. Number of drivers in each group who reacted to the changing speed limit in Event H either by a reduction of speed or by an absent acceleration before the next speed sign that set back the speed limit.

4.6.6 Drivers' self-reported awareness of driving-related information

Additionally to the objective measures of drivers' SA (drivers' reactions to the critical driving situations in Session 2), drivers were also asked to rate their awareness of different driving-related information on 7-point rating scales with the endpoints "not aware at all" and "completely aware". Subjective measures of SA were collected for both experimental sessions.

Figure 88 shows drivers' awareness ratings for different driving-related information in Session 1 and Session 2 as a function of the level of lane keeping assistance (for group means and standard deviations refer to Table L1 in Appendix L). As can be seen in Figure 88, drivers reported to have had generally a medium to high awareness of driving-related information. Also, there is a trend in the data that drivers were more aware of information that directly concerned their interaction with other road users (things that were happening on the road, presence of other road users) than of information that was relevant for vehicle control, but mostly unrelated to other traffic (position in the driving lane, road markings, and course of the road). Furthermore, it becomes evident in Figure 88 that HC drivers generally reported a lower awareness of driving-related information, especially in Session 1. Also, there was an increase in HC drivers' awareness from Session 1 to Session 2, with their ratings mostly approaching those of LDW drivers and NA drivers in Session 2.

Because the data was found to be not normally distributed (including several extreme values for each item), drivers' ratings on the different items were analysed with nonparametric analyses of variance (Kruskal-Wallis *H*) based on aligned ranks (Bortz et al., 2008). The results of these ANOVAs are summarised in Table 24.

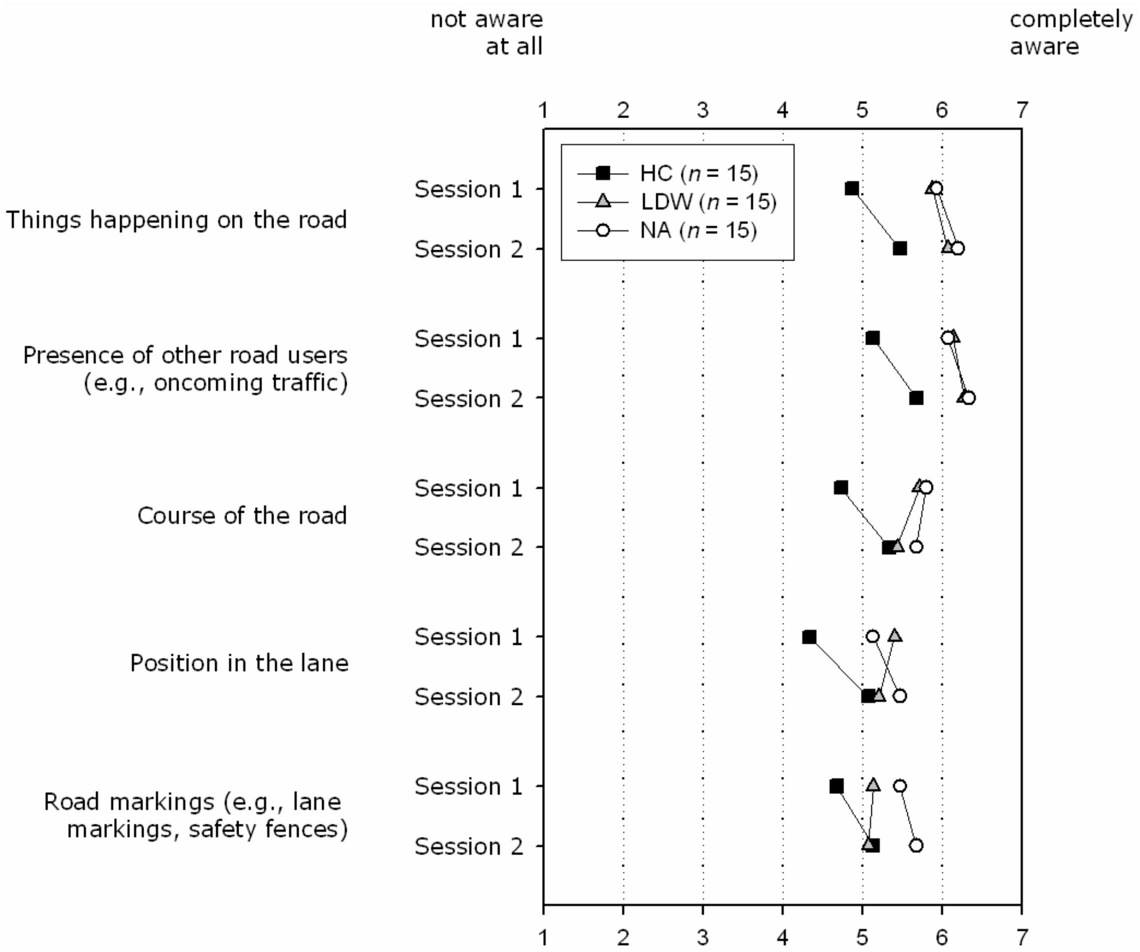


Figure 88. Drivers' awareness ratings for different types of driving-related information in Session 1 and Session 2 as a function of the level of lane keeping assistance.

These nonparametric ANOVAs revealed a significant main effect of the level of lane keeping assistance for the first two items: drivers' awareness of things happening on the road and drivers' awareness of the presence of other road users (see the ANOVAs for the first two awareness items in Table 24).

Post-hoc tests (Mann-Whitney *U*) that contrasted the HC group with the LDW and the NA group revealed that HC drivers (mean rank = 15.63) reported to be significantly less aware than LDW drivers (mean rank = 26.17) and NA drivers (mean rank = 27.20) of things happening on the road, $Z = -2.28$, $p = .02$ (two-tailed) and $Z = -2.44$, $p = .01$ (two-tailed), respectively. Also, HC drivers (mean rank = 16.37) reported to be less aware of the presence of other road users than LDW drivers (mean rank = 26.50) and NA drivers (mean rank = 26.13), although the pairwise contrasts turned out to be not significant after applying Bonferroni's correction for multiple comparisons, $Z = -2.12$, $p = .03$ (for the contrast between HC drivers and LDW drivers), and $Z = -2.09$, $p = .04$ (for the contrast between HC drivers and NA drivers).

The analyses of variance also revealed a significant main effect of session for the same two items, indicating that drivers were significantly less aware of things happening on the road in Session 1 (mean rank = 39.27) compared to Session 2 (mean rank = 51.73), $Z = -2.27$, $p = .02$ (two-tailed). Similarly, drivers' awareness of the presence of other road users significantly increased from Session 1 (mean rank = 39.09) to Session 2 (mean rank = 51.91), $Z = -2.34$, $p = .02$ (two-tailed).

For none of the other awareness items did the interaction between the level of lane keeping assistance and session become significant. There was also no significant main effect of the level of lane keeping assistance or of session for any other awareness item except for the two items discussed above.

Table 24. Nonparametric analyses of variance for awareness items.

Effect	*df*	*H*	*p*
Things happening on the road			
Level of lane keeping assistance (LoA)	2	7.60*	.02
Session (Se)	1	5.16*	.02
Se x LoA	2	0.05	.97
Presence of other road users			
Level of lane keeping assistance (LoA)	2	6.06*	.05
Session (Se)	1	5.47*	.02
Se x LoA	2	0.88	.64
Course of the road			
Level of lane keeping assistance (LoA)	2	2.65	.26
Session (Se)	1	1.40	.24
Se x LoA	2	2.38	.30
Position in the lane			
Level of lane keeping assistance (LoA)	2	4.43	.12
Session (Se)	1	3.47	.06
Se x LoA	2	5.90	.05
Road markings			
Level of lane keeping assistance (LoA)	2	4.50	.11
Session (Se)	1	0.47	.49
Se x LoA	2	0.16	.92

$^{*}p < .05$.

4.6.7 Summary and discussion

Situation awareness was the second major concept (besides drivers' reliance on the lane keeping assistance systems) that was referred to in this study in order to investigate the effect of the increasing automation of the lateral control task on drivers' level of active engagement in the driving task. Different types of critical driving situations were designed in order to study changes in drivers' cognitive processes involved in maintaining a current and comprehensive mental representation of the actual driving situation. The driving situations differed in terms of the processing demands for information that was either relevant for lateral or for longitudinal control of the vehicle; and also in terms of the temporal demands for processing of the visual cues and for response planning and execution.

It was assumed that with increasing automation of the lateral control task drivers allocate fewer attentional resources to the processing of visual information relevant for lateral control of the vehicle. This was hypothesised to result in failures in bottom-up attentional control in situations that demanded drivers' reaction primarily on the lateral control level. Those failures in bottom-up control of attention should especially become apparent in situations that placed high demands on higher-level cognitive processes (anticipation of the future behaviour of other road users – corresponding to the *lateral anticipatory* situation category) and that demanded a relatively fast reaction of drivers in order to avoid potential collisions with other traffic (*lateral unforeseeable* situation category). Measures of the latency and magnitude of drivers' reactions to the critical driving situations served as the main dependent variables and as indicators for drivers' SA.

The results did not provide clear support for this hypothesis. There were no systematic differences between the three groups of drivers in terms of the timeliness (response latency) and abruptness (response magnitude) of their reaction to these aforementioned two types of situations, such that HC drivers showed the latest and most abrupt reactions to these situations and NA drivers reacted earliest and least abruptly to these situations. According to the predictions, HC drivers reacted latest to the lateral anticipatory situation (Event G), indicated by a higher proportion of drivers in this group who started to react to this situation at a TTC of less than 4.5 s with the oncoming traffic. However, contrary to the predictions, LDW drivers and not NA drivers reacted earliest to Event G. Those differences between groups were however small and not significant.

In the lateral unforeseeable situation (Event C), a considerable difference was found between the two subgroups of HC drivers that differed significantly in their reliance on the HC system. HC drivers who did not rely on the HC system at all ('HC short' subgroup) were found to react almost 1 s earlier (in terms of TTC with the oncoming traffic) to Event C than HC drivers who relied to a large extent on the HC system ('HC long' subgroup). Furthermore, drivers who relied to a large extent on the HC system together with LDW drivers and NA drivers did not differ much in terms of their response latency to Event C. These differences between groups were marginally significant. Moreover, the groups were found to differ in the time response of their reaction to Event C, indicating differences in the groups' processing speed of the visual cues. Although drivers in the 'HC short' subgroup reacted almost 1 s earlier to Event C than drivers in the 'HC long' subgroup, both groups took a considerably longer amount of time (about 1.5 s) to apply the brake after having started to release the accelerator pedal compared to LDW and NA drivers (who applied the brake about 1 s after having started to release the accelerator pedal). On the other hand, drivers in the 'HC long' subgroup were found to reach their minimum speed significantly faster after having started to release the accelerator pedal than drivers in the 'HC short' subgroup, and also considerably faster than LDW drivers and NA drivers. This pattern of results can be interpreted as a sign for the 'HC long' subgroup's deficient processing of the visual information. Drivers in the 'HC long' subgroup did not only show a delayed reaction to Event C (similarly to LDW and NA drivers), but they apparently needed also a substantially longer amount of time to interpret the relevant visual information and to select an appropriate response (braking), unlike LDW and NA drivers. Once drivers in the 'HC long' subgroup initiated their reaction, they completed it considerably faster and in a less smooth way than the other groups of drivers. The two HC subgroups' differences in the latency and time response of their reaction to Event C could however not be attributed to their differences in visual attention allocation strategies: Drivers' total glance time to the Arrows Task display before they started to react to Event C as well as their mean single display glance duration were found to be uncorrelated with the TTC at the initial point of drivers' reactions to Event C as well as with the time between drivers' release of the accelerator pedal and their application of the brake.

The two subgroups of HC drivers did however not differ in their reaction to any other of the critical driving situations. All in all, the results for drivers' reactions to the seven safety-critical driving situations indicate that unexpected situations that demand a definite and clear response from drivers, i.e. a response unambiguous in nature, appear to be best suited to investigate potential differences in drivers' cognitive processes relevant for achieving and maintaining SA. In this study, the two unforeseeable situations most closely fulfilled this requirement. For these two situations, the largest differences between the experimental groups in terms of their reaction characteristics were observed. Whereas in Event C especially a difference between the 'HC short' subgroup and the other groups of drivers were found, Event D uncovered a difference between drivers who drove without and with lane keeping assistance. More specifically, NA drivers were found to react earlier and less abruptly to the suddenly braking vehicle in front of them in Event D than HC drivers and LDW drivers. At current, these results do not provide sufficient and consistent evidence for the hypothesised changes in drivers' cognitive processes as a function of the increasing automation of the lateral control task. No or only small and unsystematic differences between the three groups of driver were found in driving situations that left drivers a rather large time horizon for their reaction, and that were less constrained in the way how drivers could react in order to avoid a potential collision (e.g., by reducing speed and/or drawing aside).

Different explanations may account for the finding that there was no clear effect of the level of lane keeping assistance on drivers' reactions to the critical driving situations. First, it can be suggested that the hypothesised changes in drivers' cognitive processes may not be observable after such a comparably short interaction with the lane keeping assistance systems. Indeed it seems plausible to assume that changes in the way how the driving task is performed do not occur before the driving task "has actually changed", i.e. not before drivers have integrated the systems' functionality in their performance of the driving task. Thus, after about 50 min of driving, at best effects of the learning process of how to drive with the lane keeping assistance systems could probably be observed. Especially drivers driving with the HC system needed to accommodate to a car that "behaved" completely different in terms of lateral control. Drivers' evaluation of the lane keeping assistance systems in terms of usability had also revealed that drivers found the HC system to be significantly more difficult to drive with than the LDW system, and that they needed significantly more time to learn how to drive with the HC system compared to the LDW system. Although the chance for observing an effect of the level of lane keeping assistance was increased by investigating drivers' reactions to the critical driving situations under resource-intensive dual-task conditions, a much longer experimental duration is probably required for studying changes in drivers' cognitive processes in response to the increasing automation of the driving task. On the other hand it is possible that the increased density of critical driving scenarios in Session 2 resulted in a general higher "awareness" or responsiveness of drivers in the sense that they had a higher preparedness "to react" based on their experience that they were required to do so at irregular intervals. Thus, although the driving situations occurred unexpected for drivers and were different in nature, there might have been a shift in drivers' attentional focus making it impossible to observe potential automation-induced changes in drivers' SA.

Second, there was a large variation within groups found for any of the behavioural measures. On the one hand this may be attributed to the "nature" of the driving task as a complex activity which is composed of a number of subtasks requiring various perceptual, cognitive and manual skills (Groeger, 2000). Furthermore, driving in itself may be less constrained in its "execution" than other tasks, permitting for higher degrees of freedom in its performance. Also, driving is highly influenced by individual driver characteristics, such as personality traits, attitudes and motives, and cognitive abilities. Thus, in order to detect differences in drivers' reaction attributable to varying levels of automation of the lateral control task much larger sample sizes may be required. On the other hand, the large variation within groups certainly resulted from the fact that only one data point was available for each driver for each of the critical driving situations. It must however be noted that drivers' reactions when they *first* encountered the critical situations was the variable of interest based on the assumption that automation-induced changes in drivers' cognitive processes could only be observed when drivers were unprepared to react. Thus, it was assumed that drivers' repeated reactions to the same driving scenarios would rather represent a measure of drivers' recognition performance than of their SA. Another explanation for the large variation within groups might be that the critical driving situations were "not critical enough" such that drivers could react only in one possible way in order to avoid a collision – a reaction which could then be evaluated in terms of its time characteristics and appropriateness. Rather, the situations were designed in a way that left drivers some degrees of freedom in their reaction, for different reasons. First, the initial point of drivers' reaction, and not drivers' performance in general, was the main interesting variable because it was assumed to provide insight into the cognitive processes necessary to build up a current and comprehensive mental representation of the actual situation. Thus, the initial point of drivers' reaction was expected to be indicative for the quality of drivers' situation perception and comprehension, such that a delayed reaction would imply deficiencies in the cognitive processes necessary to instantiate (activate) specific types of information in the situation model. Second, the critical situations were on purpose *not* designed to force drivers into collisions, because this experience was assumed to lead to an increase in stress and anxiety in drivers, leaving all subsequent measurements of drivers' reactions unusable for the purpose of research.

A third explanation for the fact that no clear effect of the level of lane keeping assistance on drivers' reactions to the critical driving situations was found may be that the initial point of drivers' reaction may have been an insensitive measure for the hypothesised changes in cognitive processes as a function of the increasing automation of the lateral control task. The initial point of drivers' reaction was used as an indicator for drivers' SA because it was assumed to reflect most closely the point in time where drivers had *perceived* the critical driving situations. However, it is possible that drivers had actually perceived the critical situations some time before they started to react to it. Consequently, the starting point of drivers' reactions may diverge from the actual point in time at which drivers had perceived the critical situations. This may especially have been the case in situations that left drivers a comparably large amount of time in order to process the relevant visual information and to plan ahead their response. No or at most small differences between groups were found in those situations. On the other hand, drivers' reactions to the three critical situations that were categorised as "anticipatory" indicated that drivers' first observable reaction indeed reflected drivers' processing of information in terms of situation perception and comprehension. In the anticipatory situations, a proportion of drivers were found to react already *before* the situations actually evolved. Thus, in Event E for example, some drivers already applied the brake *before* the vehicle accelerated in the lay-by at the right road side in order to pull out in front of drivers. Those drivers were found to react significantly earlier to the vehicle's acceleration in the lay-by later on than drivers who did not show a reaction prior to the Event. Similarly, drivers who reacted a first time to Event G about 350 m before where they could see some oncoming vehicles overtaking the defective truck on the left road side were found to react significantly earlier to the oncoming traffic later on just before they passed the truck themselves than drivers who did not show a reaction prior to Event G. Furthermore, drivers who showed a reaction prior to Event G were found to brake significantly harder in response to the oncoming traffic just before they passed the defective truck than drivers who did not show a reaction prior to Event G. In summary, those results indicate that the initial point of drivers' reactions to the critical driving situations seemed to reflect a valid measure of drivers' situation perception and comprehension under some circumstances.

A last point concerns the extraction of the starting point of drivers' reactions from the continuous behavioural profile recorded for each driver. As already discussed in section 4.6.1.1, the criteria used to determine the initial point of drivers' reactions to the critical driving situations were suitable to detect *definite* changes in drivers' behaviour. Although the criteria were extensively tested on a large sample of drivers in different situations, and although they were found to produce valid measures of the initial point of drivers' reactions in most circumstances, it is possible that (at least in some cases) the extracted starting points of drivers' reactions did not correspond with the point in time where drivers actually initiated a response. More specifically, one may conceptualise the whole process of drivers' perception of the visual cues, the interpretation of their meaning, the selection of an appropriate response and its initiation not as a sequence of distinct points in time, but rather as an ongoing process whereof a (definite) change in behaviour is only the last instance that can be observed by help of the methods applied here.

Besides of investigating the effect of the increasing automation of the lateral control task on drivers' SA, it was tested whether drivers' reliance on the lane keeping assistance systems was related to their SA. Generally, a negative relationship is assumed between operators' reliance on automation and their SA (Sarter & Woods, 1991; Scerbo, 1996). High reliance on automation is associated with operators' less active engagement in the partly automation-controlled task. Operators were found to invest less effort in monitoring and verifying the automation's actions the more they relied on it. Similarly, it can be hypothesised that operators invest less effort in maintaining an accurate and comprehensive mental representation of the actual situation the more control they allocate to automation.

The mean duration of drivers' single display glances to the Arrows Task display was used as a measure for drivers' reliance on the lane keeping assistance systems in this study. It was found that drivers' mean single display glance duration measured during the Arrows Task where drivers encountered the respective critical driving situations was weakly negatively correlated with the THW or TTC at the initial point of drivers' reaction to them. This negative correlation was only found for the two situations that were categorised as foreseeable (Event A and Event B), and for Event F (categorised as longitudinal anticipatory). For all other types of critical driving situations, there was no relationship between drivers' reliance on the lane keeping assistance systems and their response latency to the situations. Interestingly, for Event A, B and F a highly significant negative correlation was found between drivers' total glance time to the Arrows Task display *before* drivers started to react to the Events and the THW or TTC at the initial point of drivers' reaction. Thus, the longer drivers had (in sum) glanced to the Arrows Task display when they approached the three Events, the later did they react to it. Drivers' total glance time to the Arrows Task display before they initiated their reaction was also significantly negatively correlated with the time between drivers' release of the accelerator pedal and their application of the brake (as a measure of the abruptness of the initiated manoeuvre). Thus, drivers did not only react later, but also more abruptly the longer they had (in sum) glanced to the Arrows Task display before they started to react to the three Events.

The fact that the measures of the timeliness and abruptness of drivers' reactions to Event A, B and F were strongly negatively correlated with drivers' total glance time to the Arrows Task display before drivers started to react to the Events, but only weakly correlated with the mean duration of drivers' single display glances, let suggest that reliance might *not* have been the causal factor that affected drivers' response latency and magnitude to the Events, but rather drivers' motivation to perform well on the Arrows Task. Thus, especially in foreseeable driving situations where drivers had a comparably large amount of time for processing of the visual cues and for response planning and execution, part of the drivers may have tried to respond to as many Arrows Task displays as possible before they were required to react at latest in order to avoid a collision with other road users. This interpretation of results is strengthened by the fact that the visual attention allocation measures were not correlated with the measures of the timeliness and abruptness of drivers' reactions to the critical driving situations when those occurred unexpectedly or suddenly for drivers. The finding that a correlation between measures of drivers' visual attention allocation and measures of the timeliness and abruptness of drivers' reactions was also found for Event F indicates that drivers seemed to be expecting that something might be happening when they approached Event F. Indeed, the bus on the left road side and the group of people standing in front of it were visible for drivers about 300 m before and might have presented a rather unusual scenario.

In conclusion, it can be stated that no clear and consistent evidence was found for the hypothesised negative effects of the increasing automation of the lateral control task on drivers' SA, as indicated by measures of the timeliness and abruptness of drivers' reactions to the critical driving situations that occurred unexpectedly for them in Session 2. It has however to be noted that the analyses of subjective measures of drivers' SA revealed that HC drivers reported to have been significantly less aware of things happening on the road than LDW and NA drivers. Similarly, HC drivers also reported that they were significantly less aware of the presence of other road users than LDW and NA drivers during the drive. A similar trend (although not significant) was found for drivers' self-reported awareness of other driving-related information. The current results let suggest that the HC drivers' lower self-reported awareness of driving-related information may rather be attributed to an increased involvement of conscious cognitive processes in steering control and in the interaction with the HC system than to failures in the cognitive processes necessary to achieve and maintain SA in response to the increasing automation of the lateral control task.

5 GENERAL DISCUSSION

The aim of this study was to investigate changes in drivers' active engagement in the driving task as a function of (a) the increasing automation of the lateral control task and (b) changing situational demands that affected the performance of the lane keeping assistance systems. Drivers' active engagement in the driving task was studied by referring to two established psychological concepts: drivers' reliance (on the lane keeping assistance systems) and drivers' Situation Awareness (SA).

Reliance was defined as a dynamic act of operators allocating varying levels of control to an automated system. It was assumed that drivers' reliance on a lane keeping assistance system increases the more completely the system automates the lateral control task.

The results did only partially support this hypothesis. Specifically, it was found that only a subgroup of drivers who drove with a Heading Control (HC) system (representing a high level of lane keeping assistance) relied to a considerably higher extent on the system than drivers who drove with a Lane Departure Warning (LDW) system (representing a low level of lane keeping assistance). This subgroup of HC drivers (referred to as 'HC long' subgroup) was prepared to glance away from the road scene for considerably longer periods of time during their performance of a visually demanding secondary task while driving than LDW drivers. In contrast, the other subgroup of HC drivers (referred to as 'HC short' subgroup) apparently refused to rely on the HC system at all. This second group of HC drivers was found to be even less prepared than drivers in the control group (driving without lane keeping assistance) to divert their visual attention away from the road scene, as indicated by their much shorter single glances to the secondary task display. The difference in reliance between the two subgroups of HC drivers was highly significant and remained stable over the duration of the study. Drivers driving with the LDW system did not differ from drivers in the control group (NA drivers) in their visual attention allocation strategies, indicating that they did not rely on the LDW system.

The finding that drivers apparently differed to a large extent in their willingness to allocate control to higher-level automation replicates results from previous automation studies (e.g., Lee & Moray, 1992; Lee & Moray, 1994; McFadden et al., 1998; Riley, 1996). Those studies provided evidence for large individual biases in participants' preferences for manual vs. automatic control when they were free to decide whether they wanted to perform a given task manually or to delegate it to automation. However, those studies did not go further in investigating potential determinants of individual differences in participants' reliance.

In this study, factors that may influence drivers' reliance on the lane keeping assistance systems were explicitly taken into account by observing a range of objective and subjective measures. Based on a conceptual theoretical model of humans' adaptation processes in response to automation it was suggested that operators' reliance on automation is influenced by their attitudes towards automation (i.e., their trust in it), by motivational processes (their strategic investment of mental effort) and by their energetic state (arousal and vigilance).

It turned out that different processes played a role in drivers' reliance on the lane keeping assistance systems depending on the level of assistance. The results provided strong support for the mediating influence of drivers' trust in the HC system on their reliance on it. Drivers who relied to a large extent on the HC system ('HC long' subgroup) reported a significantly higher level of trust in the system than drivers who did not rely on the HC system at all ('HC short' subgroup). Furthermore, there was a strong linear positive correlation between drivers' trust in the HC system and their reliance on it. Drivers' trust in the HC system was found to be the strongest predictor of their reliance on the HC system. As suggested by theoretical frameworks of operators' reliance on automation (e.g., Lee & See, 2004), subsequent analyses revealed that drivers' evaluation of the HC system's performance had a significant influence on their trust in it. Drivers in the 'HC long' subgroup perceived the HC system as more competent in preventing them from unintentional departures from the driving lane than drivers in the 'HC short' subgroup. Further, it turned out that drivers' evaluation of the HC system's competence was the strongest predictor of their trust in the HC system during the first experimental session, followed by drivers' perception of the comprehensibility of the HC system's behaviour.

Interestingly, drivers' trust in the LDW system did not have an influence on drivers' reliance on the LDW system. Furthermore, drivers' trust in the LDW system was found to be largely unaffected by the LDW system's performance in the first experimental session, suggesting that different mechanisms may play a role in the development of drivers' trust in a driver assistance system depending on the level of automation of the driving task.

Results did further provide evidence for the influence of energetic processes on drivers' reliance on the HC system. Drivers' arousal (as measured by the Stanford Sleepiness Scale) was found to be the second best predictor of drivers' reliance on the HC system. Together with trust, arousal accounted for 42.3% of the variance in drivers' reliance on the HC system. Thus, both a high level of trust in the HC system and a low level of arousal were significant predictors of drivers' reliance on the HC system. Again, arousal was found to have no influence on drivers' reliance on the LDW system.

No support was found for the hypothesis that drivers' reliance on the lane keeping assistance systems results (at least partially) from drivers' motivation to limit their investment of mental effort in the driving task. There was neither an indication that drivers invested less effort in the driving task, nor that they experienced lower levels of overall workload with higher levels of lane keeping assistance. Different explanations may account for this finding. First, drivers' evaluation of their interaction with the lane keeping assistance systems indicated that driving with the HC system required more conscious control than driving with the LDW system. Drivers reported that driving with the LDW system was significantly easier than driving with the HC system and that they needed significantly more time to learn how the HC system worked than to learn how the LDW system worked. Furthermore, drivers reported that they enjoyed it significantly more to drive with the LDW system than to drive with the HC system. These results let suggest that drivers' interaction with the HC system was associated with substantial cognitive costs associated with learning the HC system's functionality and how to handle the HC system's continuous steering corrections. Therefore, driving with the HC system was probably experienced as more effortful by drivers than driving with the LDW system. On the other hand, LDW drivers seemed to have invested more effort in lane keeping than NA drivers, as indicated by their superior lane keeping performance. Since the LDW system did not relieve drivers from manual control demands associated with lateral control, the better lane keeping performance of LDW drivers can be interpreted as a consequence of their greater effort to protect it from deteriorations during concurrent performance of the Arrows Task, probably in order to limit the frequency of lane departure warnings.

An unpredicted finding was that especially drivers in the 'HC long' subgroup who relied to a large extent on the HC system reported a particularly high level of workload. This speaks against the notion that those drivers' high level of reliance resulted from an adoption of an effort conservation strategy whereby they attempted to limit the expenditure of mental effort in the driving task. Results however speak rather for the assumption that drivers in the 'HC long' subgroup suffered from a vigilance decrement. Thus, drivers in the 'HC long' subgroup were found to have considerably longer response times in the Arrows Task than all other groups of drivers. At the same time, they showed (together with drivers in the 'HC short' subgroup) the worst response accuracy in the secondary task. The fact that the 'HC long' subgroup did not outperform the other groups of drivers in terms of response accuracy in the secondary task *despite* that they risked substantially longer glances to the in-vehicle display and that they took more time to respond to the secondary task argues against the notion that the 'HC long' subgroup's longer response times may have resulted from their higher motivation to perform well on the Arrows Task. Also, no significant differences between groups were found for drivers' self-reported willingness to succeed on the Arrows Task. Further support for the assumption that drivers in the 'HC long' subgroup had problems with sustained attention comes from the finding that those drivers showed a workload profile that was different from the other groups of drivers. Previous research has shown that tasks that placed high demands on sustained attention were perceived as very resource demanding by operators, and were associated with high subjective workload (Warm, Parasuraman et al., 2008). Moreover, those tasks were found to produce a consistent workload signature among subscales of the NASA-TLX with mental demands and frustration being the primary components of workload (Warm et al., 1996). In this study, drivers in the 'HC long' subgroup reported a significantly higher level of frustration than the other groups of drivers, while they reported also higher levels of mental demands, temporal demands, and effort. Also, drivers in the ‘HC long’ subgroup reported to be more dissatisfied with their own performance than drivers in the other groups. Evidence was found that the higher frustration of drivers in the ‘HC long’ subgroup partially resulted from their impression that their lane keeping performance was “not good enough”. Due to their longer eyes-off-the-road times drivers in the ‘HC long’ subgroup experienced probably more frequent and stronger interventions of the HC system which may have contributed to their higher dissatisfaction with their own lane keeping performance. All in all, these results argue for the notion that drivers in the ‘HC long’ subgroup had more difficulties to concentrate on the secondary task while driving, which in turn led to their higher levels of perceived workload, frustration, and dissatisfaction with their own performance.

No consistent evidence was found for the nature of the vigilance decrement that became apparent in the ‘HC long’ subgroup. It was hypothesised that problems with vigilance may result from drivers’ low level of arousal as a consequence of a reduction of task demands due to the increasing automation of the lateral control task. Indeed, evidence was found that drivers’ arousal decreased as a function of the increasing automation of the lateral control task. HC drivers were found to report a significantly lower level of arousal than NA drivers. Drivers in the ‘HC long’ subgroup reported the lowest level of arousal in the first experimental session. However it was also found that drivers’ arousal was not correlated with drivers’ reaction times in the Arrows Task. These findings argue for the notion that the ‘HC long’ subgroup’s vigilance decrement was not causally determined by a low level of arousal.

Another explanation for the behavioural pattern of the ‘HC long’ subgroup is that those drivers had particular problems to cope with the dual task demands. Prinzel et al. (2005) suggested that operators’ tendency to trust and to uncritically rely on automation may represent an effective workload management strategy for some operators. Since no measures of drivers’ cognitive abilities were collected in this study, it is not possible to rule out this possibility. The fact that drivers showed this behavioural pattern only when they were assisted by higher-level automation argues for the notion that individual predispositions of drivers play a more important role in influencing drivers’ adaptation to higher-level automation. Thus, drivers in the ‘HC long’ subgroup may have been less effective in time-sharing which was then further amplified by their lower level of arousal resulting from the HC system’s support.

A second major aim of this study was to investigate changes in drivers’ reliance and hypothesised mediating factors as a function of changing task demands that affected the performance of the lane keeping assistance systems. It was hypothesised that the unexpected occurrence of critical driving situations in the second experimental session (that partly represented functional limits of the lane keeping assistance systems) would lead to (a) a decrease in drivers’ reliance on the lane keeping assistance systems, (b) a decline in drivers’ trust, and to (c) an increase in drivers’ investment of mental effort in the driving task.

The results provided strong and consistent support for this hypothesis. Drivers’ reliance on the lane keeping assistance systems significantly decreased in response to their encountering of unexpected critical driving situations, as indicated by a decline in their preparedness to allocate their visual attention away from the road scene. As predicted, this decline in drivers' reliance became not only evident during drivers' very encountering of the critical driving situations, but also during "normal" driving periods in the second experimental session. Thus, a significant linear decline in the mean duration of drivers' single glances to the Arrows Task display and in the proportion of display glance time was found from "normal" driving periods in Session 1 over Session 2 to critical driving situations in Session 2. The fact that the decline in drivers' reliance became also evident in "normal" driving periods in Session 2 supports the theoretical assumption that operators' reliance on automation is guided by human appraisal processes on the costs and benefits (expected value) associated with the allocation of varying levels of control to automation in consideration of the actual task demands, the performance of the automation, and the operators' subjective state. Thus, drivers' *anticipation* of certain negative outcomes associated with high reliance on the lane keeping assistance systems (e.g., delayed reaction in critical driving situations resulting in potential collisions) was obviously the driving factor in drivers' decision to increase their engagement in the driving task in Session 2. From a theoretical point of view, the concept of *perceived risk* gains special importance in this respect (cp. Riley, 1996).

It turned further out that the decrease in drivers' reliance from Session 1 to Session 2 was accompanied by a shift in mental effort regulation strategies. Whereas drivers reported to have assigned about equal priority to the driving task and to the Arrows Task in the first experimental session, their proportion of overall effort invested in driving significantly increased in the second experimental session. There was also a significant decline in drivers' self-reported motivation to succeed in the Arrows Task in Session 2. However, the results also indicated that drivers attached considerable importance to the Arrows Task and did not regard it as a task of low priority. Thus, although drivers allocated significantly less visual attention to the Arrows Task in normal driving periods in Session 2, their performance in the Arrows Task did not deteriorate as a consequence. Apparently, drivers switched to more effective time-sharing strategies in response to the increase in driving task demands in Session 2. Additionally, drivers may have mobilised some extra effort in order to protect their performance in the secondary task. Indeed, drivers reported a significant increase in overall workload in Session 2, which was substantiated by a significant increase in workload across all six dimensions of the NASA-TLX. However, since drivers' workload ratings were not separately collected for "normal" driving periods in Session 2 and critical driving situations in Session 2, it cannot be clearly determined whether the increase in drivers' workload in Session 2 resulted from their experience of critical driving situations only.

Further evidence for a shift in drivers' mental effort regulation strategies was found in drivers' lane keeping performance data. Whereas drivers' lane keeping performance significantly improved in driving periods with concurrent Arrows Task performance from Session 1 to Session 2, their lane keeping performance under single-task conditions (driving alone) significantly deteriorated from Session 1 to Session 2. Apparently, drivers invested more effort in lane keeping in Session 2 during driving periods that were especially vulnerable to safety-critical deteriorations in driving performance because of the occasional occurrence of critical driving situations. Thus, the significant improvement in drivers' lane keeping performance in driving periods *with* concurrent Arrows Task performance in Session 2 seemed to result from drivers' active attempt to protect their driving performance from deteriorations resulting from the visual distraction by the Arrows Task. In order to achieve this, drivers reduced their eyes-off-the-road times during Arrows Task performance in Session 2, but also they probably invested more effort in lane keeping. As theories of mental effort regulation would predict (Hockey, 1997), drivers seemed to cope with the increasing task demands in Session 2 by lowering their performance standards in driving periods that were less vulnerable to deteriorations in lane keeping performance. Thus, drivers appeared to have switched to less resource-intensive strategies in driving periods *without* concurrent secondary task performance in Session 2, leading to a significant deterioration in lane keeping performance in those driving periods when compared to the first experimental session.

Drivers' secondary task performance data also indicated that drivers were not able to preserve performance standards in the Arrows Task when driving task demands increased even further, i.e. during critical driving situations. During critical driving situations drivers clearly prioritised the driving task over the Arrows Task, which resulted in a significant impairment in drivers' Arrows Task performance. The results further indicated that this shift in task priorities was more pronounced for HC drivers, as predicted. HC drivers were found to stop continuing performance of the Arrows Task to a larger extent during critical driving situations, as indicated by a higher percentage of missing responses compared to LDW and NA drivers.

Despite that drivers' reliance on the lane keeping assistance significantly decreased in response to the occurrence of critical driving situations in Session 2, this decrease was not mirrored by a similar decline in drivers' trust in the lane keeping assistance systems. Although drivers in the 'HC long' subgroup reported a slight decline in their trust in the HC system in Session 2, trust did not change from Session 1 to Session 2 for drivers in the 'HC short' subgroup. LDW drivers reported even a slight increase in their trust in the LDW system from Session 1 to Session 2. However, drivers' experience of functional limits of the lane keeping assistance systems was found to affect drivers' perception of the systems' properties. The fact that the lane keeping assistance systems offered inadequate support in some of the critical driving situations (in which drivers were required to react on the lateral control level in order to avoid potential collisions with other traffic) resulted in a significant reduction in drivers' perception of the comprehensibility of the lane keeping assistance systems' actions. The systems' competence (in preventing drivers from unintentional departures from the driving lane) was however not found to be markedly affected by drivers' experience of functional system limits. Only drivers in the 'HC long' subgroup (who relied to a large extent on the HC system) perceived the HC system as much less competent in Session 2 compared to Session 1. The finding that drivers' experience of functional system limits in Session 2 affected drivers' perception of the systems' properties, but not their trust in the lane keeping assistance systems, argues for the notion that trust seemed to represent a rather stable attitude of drivers towards the lane keeping assistance systems. Support for this notion was also found by the fact that drivers' trust in the systems in Session 1 turned out to be the single best predictor of their trust in the systems in Session 2. It must however be noted that subjective measures of trust were only collected as overall scores after the first and the second experimental session. Previous studies found that participants' trust in automation was susceptible to momentary fluctuations in the automation's performance when subjective measures of trust were collected continuously during the study (Lee & Moray, 1992, 1994; Muir & Moray, 1996).

It was further assumed that the experience of critical driving situations and of functional system limits would mostly affect HC drivers, based on the HC system's larger intervention in driving. Thus, it was hypothesised that the change in drivers' reliance and in mediating factors from Session 1 to Session 2 would be most pronounced for HC drivers, and least pronounced for NA drivers. Results provided no substantial support for this prediction. Drivers in the 'HC long' subgroup who relied to a large extent on the HC system showed a significantly larger decrease in reliance than LDW drivers from Session 1 to Session 2. However, they still relied to a larger extent on the system in Session 2 compared to the other groups of drivers. Thus, initial differences between groups were found to remain mainly stable across experimental sessions – a result that was replicated for all other measures. Thus, there was no indication that the occurrence of critical driving situations affected drivers differently depending on the level of automation of the lateral control task.

5.1 Strengths and Limitations of this Study

A major strength of this research is that changes in drivers' active engagement in the driving task were studied in relation with human adaptation processes in response to the increasing automation of the driving task. Based on an extensive review of the automation literature, a conceptual theoretical framework was developed that linked changes in operators' active task engagement to changes in human cognitive, motivational and energetic processes in response to automation. In order to determine the influence of these processes, a range of objective and subjective measures were collected. Unlike to many previous studies on automation especially in the area of road traffic, an attempt was made to uncover the *nature* of human responses to automation that were hypothesised to contribute to the out-of-the-loop performance problem under automation. Through this approach, it was possible to yield empirical evidence for the hypothesised mediating influence of drivers' attitudes (trust) and of energetic processes (arousal) on their reliance on the lane keeping assistance systems. Furthermore, evidence was found that these mediating variables had a greater influence on drivers' reliance on higher-level automation.

A further strength of this study is that it used process-oriented measures of the degree to which drivers were prepared to allocate control to the lane keeping assistance systems (reliance) as well as of drivers' SA. The results demonstrated that the mean single display glance duration used as measure for drivers' reliance was sensitive to changes in task demands as well as to individual differences in drivers' reliance on the HC system. At the same time, the results provided empirical evidence for the theoretically assumed dynamic nature of the concept of reliance on automation. Thus, rather than a discrete decision of operators to either engage or to disengage automation, reliance was found to be a graded process whereby operators dynamically allocate varying levels of control to a system based on the actual performance of the automation, the situational demands (e.g., severity of performance consequences), and the operators' subjective state (attitudes, energetic processes and motivational factors).

A third strength of the present study is that the use of an advanced moving-base driving simulator combined the advantages of high experimental control and of a comparably high degree of external validity. Thus, it was possible to study the effect of increasing automation of the driving task on drivers' reliance and SA under highly controlled and standardised experimental conditions, while still preserving a comparably high degree of realism regarding haptic, visual and auditory experiences while driving.

A major shortcoming of the current study is the relatively short experimental duration. As the results on drivers' mental workload and on drivers' evaluation of their interaction with the lane keeping assistance systems indicated, especially HC drivers seemed to have invested a considerable amount of conscious cognitive effort in their interaction with the system. Thus, drivers' interaction with the HC system apparently required from drivers substantial conscious control in learning the system's functionality and how to deal with its continuous interventions in driving. The effects found in this study therefore seem to rather represent short-term effects of *drivers' initial interaction* with the driver assistance systems than long-term adaptation effects in response to the automation of the driving task. Especially the predicted changes in drivers' cognitive processes may not have been observable after the short duration of the study. As discussed in chapter 4.6.7, no consistent and clear effect of the level of lane keeping assistance was found on drivers' reactions to the critical driving situations in Session 2 indicative drivers' impaired SA. Indeed it can be assumed that changes in cognitive processes in response to automation (attributable to changes in the task structure) cannot occur before the driving task has actually "changed", i.e. not before drivers started to integrate the systems' functionality in their performance of the driving task. This "integration phase" had apparently not yet been reached by drivers in this study. A similar explanation may hold for the fact that no differences were found for drivers' strategic investment of mental effort in the driving task as a function of the level of lane keeping assistance. It is possible that such changes in mental effort investment also become apparent not before drivers obtained sufficient experiences with a system's behaviour and before they developed an elaborated mental model of its functioning that guides their subsequent interaction with it.

Another limitation of this study is that it does not allow for the interpretation of causal relationships between variables. It is assumed in the conceptual theoretical framework (see chapter 2.2) that reliance on automation is not only influenced by human adaptation processes in response to automation, but also that the allocation of a certain amount of control to automation feeds back onto energetic, motivational, and cognitive processes. For example, low reliance on automation coupled with the investment of substantial mental resources in system monitoring will over time likely affect the operators' energetic state (decline in vigilance) and motivational processes (switching to effort conservation strategies). Likewise, high reliance on automation may lead to a decrease in arousal, an increase in operators' trust and complacency (especially when reliance does not lead to negative performance consequences), and an increase in the operators' tendency to adopt strategies of mental effort conservation. In this study, it was investigated whether the hypothesised changes in drivers' reliance as a function of the increasing automation of the lateral control task and as a function of changing task demands were *accompanied* by changes in drivers' attitudes and energetic and motivational processes that were assumed to influence drivers' reliance. Thus, for instance, the joint occurrence of a high level of reliance and a low level of arousal does not tell anything about the causal relationship between these two variables. A low level of arousal may have resulted in a high level of reliance, or vice versa.

Finally, it is not clear whether the observed differences in drivers' reliance on the lane keeping assistance systems would also become apparent under less constrained conditions and in real traffic. Prolonged eyes-off-the-road times represent a serious safety-concern in road traffic and are shown to contribute to accident risk (Horrey & Wickens, 2007; Wierwille & Tijerina, 1998). It is therefore of great importance to know whether drivers would be similarly prepared to allocate their visual attention away from the road scene under real traffic conditions when they are not obliged to do so (for example in order to perform a secondary task while driving as in this study). Especially drivers in the 'HC long' subgroup risked extremely long glances away from the road scene in this study. Results on drivers' evaluation of the supposed effects of the lane keeping assistance systems on driver behaviour and safety yielded some evidence that the effects found in this study may also generalise to real traffic conditions. Thus, drivers in the 'HC long' subgroup reported that the system would encourage them to significantly greater extent to pay more attention to things unrelated to driving compared to drivers in the 'HC short' group and LDW drivers. Also, drivers in the 'HC long' subgroup reported that the HC system would encourage them to a greater extent to drive riskier compared to drivers in the 'HC short' subgroup.

5.2 Outlook and Directions for future Research

5.2.1 Taking into account individual driver characteristics

Individual differences in drivers' reliance on higher-level lane keeping assistance have emerged as a central outcome of this study. This points to the importance of taking into account individual driver variables in future studies on drivers' adaptation to the increasing automation of the driving task. The results of this study indicate that drivers' propensity to trust and to rely on driver assistance systems seems to have an important influence on their preparedness to allocate control to a system. In this study, differences in drivers' reliance on the HC system became apparent after a 10 min familiarisation period with the system. Efforts have to be made to develop instruments or methods that reliably assess drivers' propensity to rely on driver assistance systems. A rather general scale assessing operators' complacency potential was proposed by Singh, Molloy, and Parasuraman (1993).

It was also found that self-reported driving style was able to account for a significant proportion of the variance in drivers' trust in the lane keeping assistance systems. Moreover, drivers' trust in the HC system and drivers' trust in the LDW system could be predicted by qualitatively different patterns of driving style. HC drivers were found to have greater trust in the HC system the more they described themselves as an anxious, emotional, and nervous driver (vs. as a courageous, calm, and confident driver). On the other hand, LDW drivers were found to have greater trust in the LDW system the more they described themselves as an unobservant and impulsive driver (vs. as an observant and fore-sighted driver). Thus, it is possible that different mechanism play a role in the development of drivers' trust in driver assistance systems dependent on the degree of their intervention in driving. More research is needed to provide further evidence for this assumption.

The results of this study also suggest considering individual differences in drivers' cognitive abilities to a higher extent in future research. Drivers' cognitive abilities may especially play a role when task demands are high (e.g., multiple-task situations) or when energetic processes (arousal and vigilance) are the focus of interest.

5.2.2 Long-term adaptation studies

The results of this study should be complemented by long-term studies that observe potential changes in drivers' engagement in the driving task over longer periods of drivers' interaction with driver assistance systems. Particularly changes in drivers' cognitive processes (SA) and in drivers' strategic investment of mental effort in the driving task may not be observable before a certain amount of time of drivers' interaction with the systems. Furthermore, long-term studies may shed light on the time dynamics and the relative influence of cognitive, motivational, and energetic processes on drivers' adaptation to driver assistance systems over time.

All in all, this study represented one of the first attempts to comprehensively investigate potential changes in drivers' active engagement in the driving task as a function of the increasing automation of the driving task. Drivers' reliance on the lane keeping assistance systems emerged as a useful concept in studying those changes. Future research is needed to extend the findings of this study to other types of driving automation (e.g., to driver assistance systems that intervene to a higher degree in drivers' cognitive or decision-making processes while driving, such as systems that assist drivers in lane changing) and to develop methods that reliably assess changes in drivers' cognitive processes in response to the increasing automation of the driving task.

6 REFERENCES

Adams, M. J., Tenney, Y. J., & Pew, R. W. (1995). Situation Awareness and the cognitive management of complex systems. *Human Factors, 37*(1), 85-104.

Archer, J. (2000). *Fundamental traffic safety issues concerning the use of Intelligent Transport Systems* (No. CTR2000:06). Stockholm: Royal Institute of Technology Publication.

Bagheri, N., & Jamieson, G. A. (2004). Considering subjective trust and monitoring behavior in assessing automation-induced "complacency". In D. A. Vincenzi, M. Mouloua & P. A. Hancock (Eds.), *Proceeding of the Second Human Performance, Situation Awareness and Automation Conference (HPSAA II)* (Vol. 2, pp. 54-59): Lawrence Erlbaum Associates, Inc.

Bahner, J. E., Hüper, A.-D., & Manzey, D. (2008). Misuse of automated decision aids: Complacency, automation bias and the impact of training experience. *International Journal of Human-Computer Studies, 66*(9), 688-699.

Bailey, N. R., & Scerbo, M. W. (2007). Automation-induced complacency for monitoring highly reliable systems: The role of task complexity, system experience, and operator trust. *Theoretical Issues in Ergonomics Science, 8*(4), 321-348.

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods, 37*(3), 379-384.

Baumann, M., & Krems, J. F. (2007). Situation awareness and driving: A cognitive model. In P. C. Cacciabue (Ed.), *Modelling driver behaviour in automotive environments. Critical issues in driver interactions with Intelligent Transport Systems* (pp. 253-265). London: Springer.

Baumann, M., Petzoldt, T., Groenewoud, C., Hogema, J., & Krems, J. F. (2008). The effects of cognitive tasks on predicting events in traffic. In C. Brusque (Ed.), *Proceedings of the European Conference on human centred Design for Intelligent Transport Systems* (pp. 3-11). Lyon: Humanist Publications.

Beck, H. P., Dzindolet, M. T., & Pierce, L. G. (2002). Operators' automation usage decisions and the sources of misuse and disuse. In E. Salas (Ed.), *Advances in human performance and cognitive engineering research: Automation* (Vol. 2, pp. 37-78): Elsevier Science Ltd.

Bliss, J. P., & Acton, S. A. (2003). Alarm mistrust in automobiles: How collision alarm reliability affects driving. *Applied Ergonomics, 34*(6), 499.

Bliss, J. P., Gilson, R. D., & Deaton, J. E. (1995). Human probability matching behaviour in response to alarms of varying reliability. *Ergonomics, 38*(11), 2300-2312.

Bortz, J., Lienert, G. A., & Boehnke, K. (2008). *Verteilungsfreie Methoden in der Biostatistik* (3rd ed.). Heidelberg: Springer.

Brookhuis, K. A., & De Waard, D. (2005). ADAS` acceptance and effects on behaviour: The consequences of automation. In G. Underwood (Ed.), *Traffic and transport psychology: Theory and application. Proceedings of the ICTTP 2004* (pp. 273-278): Elsevier Ltd.

Brown, C. M., & Noy, Y. I. (2004). Behavioural adaptation to in-vehicle safety measures: Past ideas and future directions. In T. Rothengatter & R. Huguenin (Eds.), *Traffic and Transport Psychology: Theory and Applications (Proceedings of the ICTTP 2000)* (pp. 25-46): Elsevier Ltd.

Buld, S., & Krüger, H.-P. (2002). *EMPHASIS (Effort-Management und Performance-Handling in sicherheitsrelevanten Situationen): Wirkungen von Assistenz und Automation auf Fahrerzustand und Fahrsicherheit* (No. 19 S 9812 7: Bundesministerium für Bildung und Forschung). Würzburg: Interdisziplinäres Zentrum für Verkehrswissenschaften an der Universität Würzburg.

Byers, J. C., Bittner, A. C., & Hill, A. G. (1989). Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? In A. Mital (Ed.), *Advances in Industrial Ergonomics and Safety* (Vol. I, pp. 481-485). London: Taylor & Francis.

Byrne, E. A., & Parasuraman, R. (1996). Psychophysiology and adaptive automation. *Biological Psychology, 42*(3), 249-268.

Cacciabue, P. C. (2006). Editorial: CTW special issue on human-centred design in automotive systems. *Cognition, Technology & Work, 8*(3), 159-160.

Cacciabue, P. C., & Martinetto, M. (2004). Driving support and user centred design approach - The case of the EUCLIDE anti-collision system. *IEEE International Conference on Systems, Man and Cybernetics*, 6465-6471.

Carsten, O. M. J. (2005). Mind over matter: Who's controlling the vehicle and how do we know. In G. Underwood (Ed.), *Traffic and transport psychology: Theory and application. Proceedings of the ICTTP 2004* (pp. 231-242): Elsevier Ltd.

Carsten, O. M. J., Merat, N., Janssen, W. H., Johannson, E., Fowkes, M., & Brookhuis, K. A. (2005). HASTE Final Report. Retrieved December 19, 2008, from http://www.its.leeds.ac.uk/projects/haste/downloads/Haste_D6.pdf

Carsten, O. M. J., & Nilsson, L. (2001). Safety assessment of driver assistance systems. *European Journal of Transport and Infrastructure Research, 1*(3), 225 - 243.

Chin, J. P., Diehl, V. A., & Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. Washington, D.C., United States, May 15-19.

Cohen, M. S., Parasuraman, R., & Freeman, J. T. (1999). *Trust in decisions aids - A model and its training implications* (No. Tech. Report USAATCOM TR 97-D-4). Arlington, VA: Cognitive Technologies.

Davies, D. R., & Parasuraman, R. (1982). *The psychology of vigilance*. London: Academic Press.

De Vries, P., Midden, C., & Bouwhuis, D. (2003). The effects of errors on system trust, self-confidence, and the allocation of control in route planning. *International Journal of Human-Computer Studies, 58*, 719-735.

De Waard, D. (1996). *The measurement of drivers' mental workload.* Doctoral Dissertation, Traffic Research Centre, University of Groningen, Haren, The Netherlands.

De Waard, D., Van der Hulst, M., Hoedemaeker, M., & Brookhuis, K. A. (1999). Driver behavior in an emergency situation in the automated highway system. *Transportation Human Factors, 1*(1), 67-82.

Dekker, S., & Hollnagel, E. (2004). Human factors and folk models. *Cognition, Technology & Work, 6*, 79-86.

Desmond, P. A., & Hancock, P. A. (2001). Active and passive fatigue states. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload, and fatigue* (pp. 455-465). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Desmond, P. A., Hancock, P. A., & Monette, J. L. (1998). Fatigue and automation-induced impairments in simulated driving performance. *Transportation Research Record*(1628), 8-14.

Dixon, S. R., & Wickens, C. D. (2006). Automation reliability in unmanned aerial vehicle control: A reliance-compliance model of automation dependence. *Human Factors, 48*(3), 474-486.

Dixon, S. R., Wickens, C. D., & McCarley, J. S. (2007). On the independence of comliance and reliance: Are automation false alarms worse than misses? *Human Factors, 49*(4), 564-572.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1*(2), 170-177.

Durso, F. T., & Gronlund, S. D. (1999). Situation awareness. In F. T. Durso, R. S. Nickerson, R. W. Schvaneveldt, S. T. Dumais, D. S. Lindsay & M. T. H. Chi (Eds.), *Handbook of Applied Cognition* (pp. 284-314). New York: Wiley.

Durso, F. T., Rawson, K. A., & Girotto, S. (2007). Comprehension and situation awareness. In F. T. Durso, R. S. Nickerson, S. T. Dumais, S. Lewandowsky & T. J. Perfect (Eds.), *Handbook of applied cognition* (2 ed., pp. 163-193). Chichester, UK: Wiley.

Durso, F. T., & Sethumadhavan, A. (2008). Situation Awareness: Understanding dynamic environments. *Human Factors: Golden Anniversary Special Issue, 50*(3), 442-448.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies, 58*, 697-718.

Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, L. A. (2002). The perceived utility of human and automated aids in a visual detection task. *Human Factors, 44*(1), 79-94.

Endsley, M. R. (1995a). Measurement of situation awareness in dynamic systems. *Human Factors, 37*(1), 65-84.

Endsley, M. R. (1995b). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*(1), 32-64.

Endsley, M. R. (1996). Automation and situation awareness. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications.* (pp. 163-181). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Endsley, M. R. (2000a). Direct measurement of situation awareness: Validity and use of SAGAT. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness analysis and measurement.* (pp. 147-173). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Endsley, M. R. (2000b). Theoretical underpinnings of situation awareness: A critical review. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 3-32). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Endsley, M. R., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics, 42*(3), 462-492.

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.

Engström, J., & Hollnagel, E. (2007). A general conceptual framework for modelling behavioural effects of driver support functions. In P. C. Cacciabue (Ed.), *Modelling driver behaviour in automotive environments. Critical issues in driver interactions with Intelligent Transport Systems* (pp. 61-84). London: Springer.

Engström, J., Johansson, E., & Östlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F, 8*, 97-120.

European Commission. (2001). *White Paper - European transport policy for 2010: time to decide*. Luxembourg: Office for Official Publications of the European Communities.

European Commission. (2006). Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions on the Intelligent Car Initiative "Raising Awareness of ICT for Smarter, Safer and Cleaner Vehicles". Retrieved October 10, 2008, from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2005:0229:FIN:EN:PDF

European Commission. (2009). *EU energy and transport in figures 2009*. Luxembourg: Office for Official Publications of the European Communities.

Ezer, N., Fisk, A. D., & Rogers, W. A. (2008). Age-related differences in reliance behavior attributable to costs within a human-decision aid system. *Human Factors, 50*(6), 853-863.

Fairclough, S. H. (2001). Mental effort regulation and the functional impairment of the driver. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload, and fatigue* (pp. 479-502). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Farrell, S., & Lewandowsky, S. (2000). A connectionist model of complacency and adaptive recovery under automation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(2), 395-410.

Flach, J. M. (1995). Situation awareness: Proceed with caution. *Human Factors, 37*(1), 149-157.

Floudas, N., Amditis, A., Keinath, A., Bengler, K., & Engeln, A. (2004). *Review and taxonomy of IVIS/ADAS applications (Del. 2.1.2 of AIDE IST-1-507674-IP).*

Furr, R. M., & Rosenthal, R. (2003). Repeated-measures contrasts for "multiple-pattern" hypotheses. *Psychological Methods, 8*(3), 275-293.

Galster, S. M., & Parasuraman, R. (2004). Task dependencies in stage-based examinations of the effects of unreliable automation. In D. A. Vincenzi, M. Mouloua & P. A. Hancock (Eds.), *Proceeding of the Second Human Performance, Situation Awareness and Automation Conference (HPSAA II)* (Vol. 2, pp. 23-27): Lawrence Erlbaum Associates, Inc.

Gao, J., & Lee, J. D. (2006). Extending the decision field theory to model operators' reliance on automation in supervisory control situations. *IEEE Transactions on Systems, Man & Cybernetics--Part A: Systems and Humans, 36*(5), 943-959.

Gelau, C., & Krems, J. F. (2009). *Zeitschrift für Arbeitswissenschaft. Sonderheft: Situation Awareness, 36*(1).

Getty, D. J., Swets, J. A., Pickett, R. M., & Gonthier, D. (1995). System operator response to warnings of danger: A laboratory investigation of the effects of the predictive value of a warning on human response time. *Journal of Experimental Psychology: Applied, 1*(1), 19-33.

Godthelp, H., Milgram, P., & Blaauw, G. J. (1984). The development of a time-related measure to describe driving strategy. *Human Factors, 26*(3), 257-268.

Goodrich, M. A., & Boer, E. R. (2003). Model-based human-centered task automation: A case study in ACC system design. *IEEE Transactions on Systems, Man, and Cybernetics--Part A: Systems and Humans, 33*(3), 325-336.

Gray, R., & Regan, D. (2000). Risky driving behaviour: A consequence of motion adaptation for visually guided motor action. *Journal of Experimental Psychology: Human Perception and Performance, 26*(6), 1721-1732.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Grier, R. A., Warm, J. S., Dember, W. N., Matthews, G., Galinsky, T. L., Szalma, J. L., et al. (2003). The vigilance decrement reflects limitations in effortful Attention, not mindlessness. *Human Factors, 45*(3), 349-359.

Griffiths, P. G., & Gillespie, R. B. (2005). Sharing control between humans and automation using haptic interface: Primary and secondary task performance benefits. *Human Factors, 47*(3), 574-590.

Groeger, J. A. (2000). *Understanding driving: Applying cognitive psychology to a complex everyday task*. Hove: Psychology Press.

Hancock, P. A., & Parasuraman, R. (1992). Human factors and safety in the design of intelligent vehicle-highway systems (IVHS). *Journal of Safety Research, 23*(4), 181-198.

Hancock, P. A., Parasuraman, R., & Byrne, E. A. (1996). Driver-centered issues in advanced automation for motor vehicles. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 337-364). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31*(5), 519-537.

Hargutt, V., & Krüger, H. P. (2001). The danger of underload? Effort management in the driving simulation (Die Gefahr der Unterforderung? Selbstregulation der Aktivierung in der Fahrsimulation). *VDI-Berichte, 1613*, 33-48.

Harris, W. C., Hancock, P. A., Arthur, E. J., & Caird, J. K. (1995). Performance, workload, and fatigue changes associated with automation. *International Journal of Aviation Psychology, 5*(2), 169-185.

Hart, S. G., & Staveland, L. E. (1988). Development of a NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: North-Holland.

Helton, W. S., Hollander, T. D., Warm, J. S., Matthews, G., Dember, W. N., Wallaart, M., et al. (2005). Signal regularity and the mindlessness model of vigilance. *British Journal of Psychology, 96*, 249-261.

Hitchcock, E. M., Warm, J. S., Matthews, G., Dember, W. N., Shear, P. K., Tripp, L. D., et al. (2003). Automation cueing modulates cerebral blood flow and vigilance in a simulated air traffic control task. *Theoretical Issues in Ergonomics Science, 4*(1-2), 89-112.

Hoc, J.-M., Mars, F., Milleville-Pennel, I., Jolly, É., Netto, M., & Blosseville, J.-M. (2006). Human-machine cooperation in car driving for lateral safety: Delegation and mutual control. *Le Travail Humain, 69*(2), 153-182.

Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology, 45*, 73-93.

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W. C. (1973). Quantification of sleepiness: A new approach. *Psychophysiology, 10*, 431-436.

Hoedemaeker, M., & Brookhuis, K. A. (1998). Behavioral adaptation to driving with an adaptive cruise control (ACC). *Transportation Research Part F: Traffic Psychology and Behaviour, 1*(2), 95-106.

Hoffman, J. D., Lee, J. D., McGehee, D. V., Macias, M., & Gellatly, A. W. (2005). Visual sampling of in-vehicle text messages: Effects of number of lines, page presentation, and message control, *Transportation Research Record* (pp. 22-30).

Hollnagel, E. (2002). Cognition as control: A pragmatic approach to the modelling of joint cognitive systems. Retrieved December 17, 2008, from http://www.ida.liu.se/~eriho/images/IEEE_SMC_Cognition_as_control.pdf

Hollnagel, E., & Bye, A. (2000). Principles for modelling function allocation. *International Journal of Human-Computer Studies, 52*(2), 253-265.

Hollnagel, E., & Woods, D. D. (2005). *Joint cognitive systems: Foundations of cognitive systems engineering*. Boca Raton, FL: Taylor and Francis/CRC Press.

Horrey, W. J., & Wickens, C. D. (2007). In-vehicle glance duration: Distributions, tails and a model of crash risk. *Transportation Research Record, 2018*(22-28).

Humphreys, M. S., & Revelle, W. (1984). Personality, motivation, and performance: A theory of the relationship between individual differences and information processing. *Psychological Review, 91*(2), 153-184.

Janssen, W., Wierda, M., & Van der Horst, R. (1995). Automation and the future of driver behavior. *Safety Science, 19*(2), 237-244.

Jonah, B. A. (1997). Sensation seeking and risky driving: A review and synthesis of the literature. *Accident Analysis & Prevention, 29*(5), 651-665.

Kaber, D. B., & Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science, 5*(2), 113-153.

Kaber, D. B., Perry, C. M., Segall, N., McClernon, C. K., & Prinzel, L. J. (2006). Situation awareness implications of adaptive automation for information processing in an air traffic control-related task. *International Journal of Industrial Ergonomics, 36*(5), 447-462.

Kahnemann, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kerr, N. L., & Bruun, S. E. (1983). Dispensability of member effort and group motivation losses: Free-rider effects. *Journal of Personality and Social Psychology, 44*, 78-94.

Kessel, C. J., & Wickens, C. D. (1982). The transfer of failure-detection skills between monitoring and controlling dynamic systems. *Human Factors, 24*(1), 49-60.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York, NY: Cambridge University Press.

Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decision making. In G. A. Klein, J. Orasanu, R. Calderwood & C. E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 138-147). Westport, CT, US: Ablex Publishing.

Lansdown, T. C. (2000). Driver visual allocation and the introduction of intelligent transport systems. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 214*(6), 645-652.

Lansdown, T. C. (2001). Causes, measures, and effects of driver visual workload. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload, and fatigue.* (pp. 351-369). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Latane, B., Williams, K. D., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology, 37*, 822-832.

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

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies, 40*, 153-184.

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

Lewandowsky, S., Mundy, M., & Tan, G. P. A. (2000). The dynamics of trust: Comparing humans to automation. *Journal of Experimental Psychology: Applied, 6*(2), 104-123.

Ma, R., & Kaber, D. B. (2005). Situation awareness and workload in driving while using adaptive cruise control and a cell phone. *International Journal of Industrial Ergonomics, 35*(10), 939-953.

Ma, R., & Kaber, D. B. (2007). Situation awareness and driving performance in a simulated navigation task. *Ergonomics, 50*(8), 1351-1364.

Madhavan, P., Wiegmann, D. A., & Lacson, F. C. (2006). Automation failures on tasks easily performed by operators undermine trust in automated aids. *Human Factors, 48*(2), 241-256.

Maltz, M., & Meyer, J. (2001). Use of warnings in an attentionally demanding detection task. *Human Factors, 43*(2), 217-226.

Maltz, M., & Shinar, D. (2007). Imperfect in-vehicle collision avoidance warning systems can aid distracted drivers. *Transportation Research Part F, 10*, 345-357.

Marchau, V. A. W. J., & Van der Heijden, R. E. C. M. (2000). Introducing advanced electronic driver support systems: An exploration of market and technological uncertainties. *Transport Reviews, 20*(4), 421-433.

Masalonis, A. J., Duley, J. A., & Parasuraman, R. (1999). Effects of manual and autopilot control on mental workload and vigilance during simulated general aviation flight. *Transportation Human Factors, 1*(2), 187-200.

Matthews, G. (2002). Towards a transactional ergonomics for driver stress and fatigue. *Theoretical Issues in Ergonomics Science, 3*(2), 195-211.

Matthews, G., & Davies, D. R. (1998). Arousal and vigilance: The role of task demands. In R. R. Hoffman, M. F. Sherrick & J. S. Warm (Eds.), *Viewing psychology as a whole: The integrative science of William N. Dember* (pp. 113-144). Washington, DC: American Psychological Association.

Matthews, G., Davies, D. R., & Lees, J. L. (1990). Arousal, extraversion, and individual differences in resource availability. *Journal of Personality and Social Psychology, 59*(1), 150-168.

Matthews, G., & Desmond, P. A. (2002). Task-induced fatigue states and simulated driving performance. *The Quarterly Journal of Experimental Psychology, 55A*(2), 659-686.

Matthews, G., Joyner, L., Gilliland, K., Campbell, S., Falconer, S., & Huggins, J. (1999). Validation of a comprehensive stress state questionnaire: Towards a state 'big three'. In I. J. Mervielde, F. Deary, F. DeFruyt & F. Ostendorf (Eds.), *Personality psychology in Europe* (pp. 335-350). Tilburg: Tilburg University Press.

McCarley, J. S., Wickens, C. D., Goh, J., & Horrey, W. J. (2002). *A computational model of attention/situation awareness.* Paper presented at the 46th Annual Meeting of the Human Factors and Ergonomics Society, Santa Monica.

McFadden, S. M., Giesbrecht, B. L., & Gula, C. A. (1998). Use of an automatic tracker as a function of its reliability. *Ergonomics, 41*(4), 512-536.

Metzger, U., & Parasuraman, R. (2005). Automation in future air traffic management: Effects of decision aid reliability on controller performance and mental workload. *Human Factors, 47*(1), 35-49.

Meyer, J. (2001). Effects of warning validity and proximity on responses to warnings. *Human Factors, 43*(4), 563-572.

Meyer, J. (2004). Conceptual issues in the study of dynamic hazard warnings. *Human Factors, 46*(2), 196-204.

Molloy, R., & Parasuraman, R. (1996). Monitoring an automated system for a single failure: Vigilance and task complexity effects. *Human Factors, 38*(2), 311-322.

Moray, N. (1986). Monitoring behavior and supervisory control. In K. R. Boff, L. Kaufman & J. P. Thomas (Eds.), *Handbook of perception and human performance. Vol. II: Cognitive processes and performance* (pp. 40/41-40/51). New York: John Wiley & Sons.

Moray, N., & Inagaki, T. (2000). Attention and complacency. *Theoretical Issues in Ergonomics Science, 1*(4), 354-365.

Moray, N., Inagaki, T., & Itoh, M. (2000). Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *Journal of Experimental Psychology: Applied, 6*(1), 44-58.

Mosier, K. L., & Skitka, L. J. (1996). Human decision makers and automated decision aids: Made for each other? In R. Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance: Theory and Application* (pp. 201-220). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies, 27*, 527-539.

Muir, B. M. (1994). Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics, 37*(11), 1905-1922.

Muir, B. M., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics, Special issue: Cognitive Ergonomics V, 39*(3), 429-460.

Nilsson, L. (2005). Automated driving does not work without the involvement of the driver. In G. Underwood (Ed.), *Traffic and transport psychology: Theory and application. Proceedings of the ICTTP 2004* (pp. 273-301): Elsevier Ltd.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research and theory* (Vol. 4, pp. 1-18). New York: Plenum Press.

Olejnik, S., & Algina, J. (2003). Generalized eta squared and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods, 8*(4), 434-447.

Östlund, J., Nilsson, L., Carsten, O. M. J., Merat, N., Jamson, H., Jamson, S., et al. (2004). *HMI and safety-related driver performance (Del. 2 of HASTE Project GRD1/2000/25361 S12.319626).*

Östlund, J., Peters, B., Thorslund, B., Engström, J., Markkula, G., Keinath, A., et al. (2005). *Driving performance assessment - Methods and metrics (Del. 2.2.5 of AIDE IST-1-507674-IP).*

Parasuraman, R. (1987). Human-computer monitoring. *Human Factors, 29*(6), 695-706.

Parasuraman, R., & Hancock, P. A. (2001). Adaptive control of mental workload. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload, and fatigue* (pp. 305-320). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Parasuraman, R., Molloy, R., Mouloua, M., & Hilburn, B. (1996). Monitoring of automated systems. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 91-115). Mahwah, NJ: Lawrence Erlbaum Associates.

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced "complacency". *The International Journal of Aviation Psychology, 3*(1), 1-23.

Parasuraman, R., & Mouloua, M. (Eds.). (1996). *Automation and human performance: Theory and applications*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors, 38*(4), 665-679.

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 of types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 30*(3), 286-297.

Parasuraman, R., & Wickens, C. D. (2008). Humans: Still vital after all these years of automation. *Human Factors: Golden Anniversary Special Issue, 50*(3), 511-520.

Prinzel, L. J., Freeman, F. G., & Prinzel, H. D. (2005). Individual differences in complacency and monitoring for automation failures. *Individual Differences Research, 3*(1), 27-49.

Rauch, N., Gradenegger, B., & Krüger, H. P. (2006). The measurement of situation awareness while driving with a secondary task (Die Erfassung von Situationsbewusstsein beim Fahren mit Nebenaufgabe). *VDI-Berichte, 1960*, 57-73.

RESPONSE 3 Code of practice for the design and evaluation of ADAS. (2006). Retrieved November 25, 2008, from http://prevent-ip.org/download/deliverables/RESPONSE3/D11.2/Response3_CoP_v3.0.pdf

Riley, V. (1996). Operator reliance on automation: Theory and data. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 19-35). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral research. A correlational approach*. New York: Cambridge University Press.

Rousseau, R., Tremblay, S., & Breton, R. (2004). Defining and modeling situation awareness: A critical review. In S. Banbury & S. Tremblay (Eds.), *A cognitive approach to situation awareness: Theory and application* (pp. 3-21). Aldershot, UK: Ashgate.

Rovira, E., McGarry, K., & Parasuraman, R. (2007). Effects of imperfect automation on decision making in a simulated command and control task. *Human Factors, 49*(1), 76-87.

Rudin-Brown, C. M., & Noy, Y. I. (2002). Investigation of behavioral adaption to lane departure warnings. *Transportation Research Record, 1803*, 30-37.

Saad, F. (2004). *Behavioural adaptations to new driver support systems: Some critical issues.* Paper presented at the IEEE International Conference on Systems, Man and Cybernetics.

Sarter, N. B., & Woods, D. D. (1991). Situation awareness: A critical but ill-defined phenomenon. *The International Journal of Aviation Psychology, 1*(1), 45-57.

Sarter, N. B., & Woods, D. D. (1992). Pilot interaction with cockpit automation: Operational experiences with the flight management system. *International Journal of Aviation Psychology, 2*(4), 303-321.

Scallen, S. F., & Hancock, P. A. (2001). Implementing adaptive function allocation. *International Journal of Aviation Psychology, 11*(2), 197-221.

Scerbo, M. W. (1996). Theoretical perspectives on adaptive automation. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 37-63). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Sedlmeier, P., & Renkewitz, F. (2008). *Forschungsmethoden und Statistik in der Psychologie*. München: Pearson Studium.

Seppelt, B. D., & Lee, J. D. (2007). Making adaptive cruise control (ACC) limits visible. *International Journal of Human-Computer Studies, 65*(3), 192-205.

Shebilske, W. L., Goettl, B. P., & Garland, D. J. (2000). Situation awareness, automaticity, and training. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 303-323). Mahwah: Lawrence Erlbaum.

Sheridan, T. B. (2006). Supervisory control. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (3rd ed., pp. 1025-1052). Hoboken, NJ, US: John Wiley & Sons Inc.

Sheridan, T. B., & Hennessy, R. T. (Eds.). (1984). *Research and modeling of supervisory control behavior. Report of a workshop*. Washington, D.C.: National Academy Press.

Sheridan, T. B., & Parasuraman, R. (2005). Human-automation interaction. *Reviews of Human Factors and Ergonomics, 1*, 89-129.

Sheridan, T. B., & Verplanck, W. L. (1978). *Human and computer control of undersea teleoperators.* MIT Man-Machine Systems Laboratory, Cambridge, MA.

Shneiderman, B., & Plaisant, C. (2005). *Designing the user interface: Strategies for effective Human-Computer Interaction.* (4 ed.). Boston, Mass.: Pearson/Addison-Wesley.

Simon, J. H. (2005). *Learning to drive with Advanced Driver Assistance Systems. Empirical studies of an online tutor and a personalised warning display on the effects of learnability and the acquisition of skill.* Doctoral dissertation, Philosophische Fakultät, Technische Universität Chemnitz.

Singh, I. L., Molloy, R., & Parasuraman, R. (1993). Automation-induced "complacency": Development of the complacency-potential rating scale. *The International Journal of Aviation Psychology, 3*(2), 111-122.

Singh, I. L., Molloy, R., & Parasuraman, R. (1997). Automation-induced monitoring inefficiency: role of display location. *International Journal of Human-Computer Studies, 46*, 17-30.

Skitka, L. J., Mosier, K., & Burdick, M. D. (2000). Accountability and automation bias. *International Journal of Human-Computer Studies, 52*(4), 701-717.

Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies, 51*(5), 991-1006.

Smith, K., & Hancock, P. A. (1995). Situation awareness is adaptive, externally directed consciousness. *Human Factors, 37*(1), 137-148.

Sorkin, R. D., & Woods, D. D. (1985). Systems with Human Monitors: A signal detection analysis. *Human-Computer Interaction, 1*(1), 49-75.

Stanton, N. A., & Young, M. S. (1998). Vehicle automation and driving performance. *Ergonomics, 41*(7), 1014-1028.

Stanton, N. A., & Young, M. S. (2005). Driver behaviour with adaptive cruise control. *Ergonomics, 48*(10), 1294-1313.

Strayer, D. L., Drews, F. A., & Johnston, W. A. (2003). Cell phone-induced failures of visual attention during simulated driving. *Journal of Experimental Psychology: Applied, 9*(1), 23-32.

Tango, F., & Montanari, R. (2006). Shaping the drivers' interaction: How the new vehicle systems match the technological requirements and the human needs. *Cognition, Technology and Work, 8*(3), 215- 226.

Tejero, P., & Chóliz, M. (2002). Driving on the motorway: The effect of alternating speed on driver's activation level and mental effort. *Ergonomics, 45*(9), 605-618.

Tenney, Y. J., & Pew, R. W. (2006). Situation awareness catches on. What? So what? Now what? In R. C. Williges (Ed.), *Reviews of human factors and ergonomics* (Vol. 2, pp. 89-129). Santa Monica, CA: Human Factors and Ergonomics Society.

Tsimhoni, O., & Green, P. (2001). *Visual demand of driving and the execution of display-intensive in-vehicle tasks.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting.

Tsimhoni, O., & Green, P. (2003). Time-sharing of a visual in-vehicle task while driving: Effects of four key constructs. In *Proceedings of the 2nd international driving symposium on human factors in driver assessment, training, and vehicle design* (pp. 113-118). Park City, Utah.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Acadamic Press.

Van Winsum, W., Brookhuis, K. A., & De Waard, D. (2000). A comparison of different ways to approximate time-to-line crossing (TLC) during car driving. *Accident Analysis & Prevention, 32*(1), 47-56.

Verwey, W. B., & Veltman, H. A. (1996). Detecting short periods of elevated workload: A comparison of nine workload assessment techniques. *Journal of Experimental Psychology: Applied, 2*(3), 270-285.

Victor, T. W., Harbluk, J. L., & Engström, J. A. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research: Part F, 8*(2), 167-190.

Walker, G. H., Stanton, N. A., & Young, M. S. (2001). Where is computing driving cars? *International Journal of Human-Computer Interaction, 13*(2), 203-229.

Warm, J. S., & Dember, W. N. (1998). Test of vigilance taxonomy. In R. R. Hoffman, M. F. Sherrick & J. S. Warm (Eds.), *Viewing psychology as a whole: The integrative science of William N. Dember* (pp. 87-112). Washington, DC: American Psychological Association.

Warm, J. S., Dember, W. N., & Hancock, P. A. (1996). Vigilance and workload in automated systems. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and Applications* (pp. 183-200). Mahwah, NJ: Lawrence Erlbaum Associates.

Warm, J. S., Matthews, G., & Finomore, V. S. (2008). Workload, stress, and vigilance. In P. A. Hancock & J. L. Szalma (Eds.), *Performance under stress* (pp. 115-141). Brookfield, VT: Ashgate.

Warm, J. S., & Parasuraman, R. (2007). Cerebral hemodynamics and vigilance. In R. Parasuraman & M. Rizzo (Eds.), *Neuroergonomics: The brain at work* (pp. 146-158). New York: Oxford University Press.

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors: Golden Anniversary Special Issue, 50*(3), 433-441.

Wickens, C. D. (2001). Workload and situation awareness. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload, and fatigue* (pp. 443-450). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Wickens, C. D. (2008). Situation Awareness: Review of Mica Endsley's 1995 articles on Situation Awareness theory and measurement. *Human Factors: Golden Anniversary Special Issue, 50*(3), 397-403.

Wickens, C. D., & Colcombe, A. (2007). Dual-task Performance consequences of imperfect alerting associated with a cockpit display of traffic information. *Human Factors, 49*(5), 839-850.

Wickens, C. D., & Dixon, S. R. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science, 8*(3), 201-212.

Wickens, C. D., & Kessel, C. (1981). Failure detection in dynamic systems. In J. Rasmussen & W. B. Rouse (Eds.), *Human detection and diagnosis of system failures. NATO Conference Series* (Vol. 15, pp. 155-169): Plenum Press.

Wiegmann, D. A., Rich, A., & Zhang, H. (2001). Automated diagnostic aids: The effects of aid reliability on users' trust and reliance. *Theoretical Issues in Ergonomics Science, 2*(4), 352-367.

Wierwille, W. W. (1993). Visual and manual demands of in-car controls and displays. In B. Peacock & W. Karwowski (Eds.), *Automotive Ergonomics* (pp. 299-320). London: Taylor & Francis Ltd.

Wierwille, W. W., & Tijerina, L. (1998). Modelling the relationship between driver in-vehicle visual demands and accident occurence. In A. G. Gale, I. Brown, C. M. Haslegrave & S. P. Taylor (Eds.), *Vision in vehicles IV* (pp. 233-243). Amsterdam: Elsevier.

Wikman, A.-S., Nieminen, T., & Summala, H. (1998). Driving experience and time-sharing during in-car tasks on roads of different width. *Ergonomics, 41*(3), 358-372.

Yeh, M., & Wickens, C. D. (2001). Display signaling in augmented reality: Effects of cue reliability and image realism on attention allocation and trust calibration. *Human Factors, 43*(3), 355-365.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit information. *Journal of Comparative and Neurological Psychology, 18*, 459-482.

Young, K. L., Regan, M. A., & Lee, J. D. (2008). Measuring the effects of driver distraction: Direct driving performance measures and methods. In M. A. Regan, J. D. Lee & K. L. Young (Eds.), *Driver distraction: Theory, effects and mitigation*. Boca Raton, FL: Taylor & Francis, CRC Press.

Young, M. S., & Stanton, N. A. (1997). Automotive automation: Investigating the impact on drivers` mental workload. *International Journal of Cognitive Ergonomics, 1*(4), 325-336.

Young, M. S., & Stanton, N. A. (2002a). Attention and automation: New perspectives on mental underload and performance. *Theoretical Issues in Ergonomics Science, 3*(2), 178-194.

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

Zhang, H., Smith, M. R. H., & Witt, G. J. (2006). Identification of real-time diagnostic measures of visual distraction with an automatic eye-tracking system. *Human Factors, 48*(4), 805-821.

Zijlstra, F. R. H. (1993). *Efficiency in work behavior. A design approach for modern tools.* Doctoral dissertation, Delft University of Technology, Delft, The Netherlands.

Zwahlen, H. T., Adams, C. C., & DeBald, D. P. (1988). Safety aspects of CRT touch panel controls in automobiles. In A. G. Gale, M. H. Freeman, C. M. Haslegrave, P. Smith & S. P. Taylor (Eds.), *Second International conference on Vision in Vehicles, Nottingham, England*. Amsterdam: Elsevier Science, North-Holland.

APPENDIX A: QUESTIONNAIRE

On the following pages we ask you to report about your subjective experiences during the preceding drive.

Please answer all the questions. Try to not think too much about the single questions, your first answer is usually the best.

All collected data will be used exclusively for scientific purposes and will be treated entirely confidential. No data will be transferred that allows for inference about the performance or behaviour of individual persons.

If you have problems with any of the questions please refer to the experimental leader.

Thank you very much for your support!

Part 1: Alertness

Below you find some descriptors of how alert or sleepy you might have been feeling during the preceding drive.

Please read them carefully and circle the number that best corresponds to the statement describing how you felt during the preceding drive.

1 Feeling active and vital.
Alert and wide awake.

2 Functioning at a high level, but not at peak.
Able to concentrate.

3 Relaxed and awake, but not at full alertness.
Responsive.

4 A little foggy, not at peak.
Let down.

5 More foggy. Beginning to lose interest in staying awake.
Slowed down.

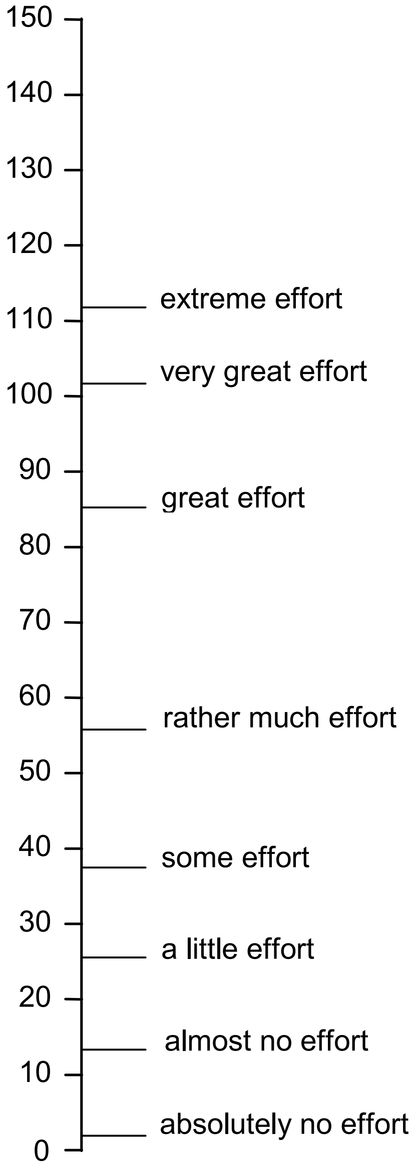
6 Very sleepy, fighting sleep, woozy.
Prefer to be lying down.

7 Almost asleep.
Lost struggle to remain awake.

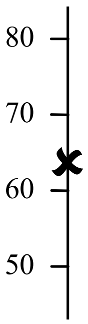
Part 2: Mental Effort

Overall Mental Effort

Please indicate, by marking the vertical axis below, how much effort it took for you to *drive and concurrently perform the Arrows Task* during the preceding drive.



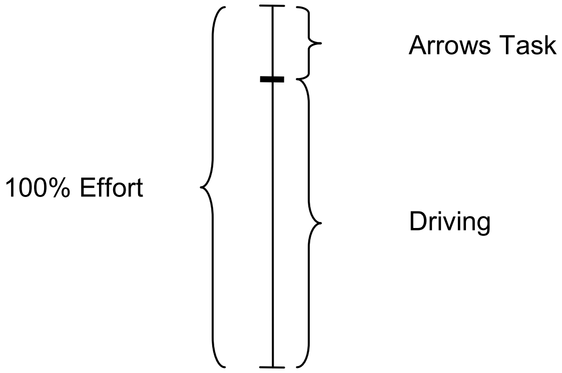
Example:



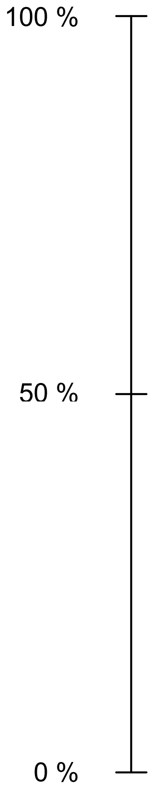
Mental Effort subdivided into Driving and Arrows Task

Now we want to know in more detail how much effort you put in driving and in the Arrows Task.
Please suppose that your overall effort during the preceding drive is represented by the length of the line below, corresponding to 100 % of your overall effort. Please indicate now, by separating the line into two parts, how much effort you put in driving (in the lower part of the line) and how much effort you put in the Arrows Task (in the upper part of the line) during the preceding drive.

Example:
If you found driving very demanding, you may have had not many resources left to put in the Arrows Task. You might subdivide your overall effort as follows:



Please mark now the line below on the position which best corresponds to the allocation of your effort during the preceding drive:



Please mark the box which applies most for you:

My driving performance while performing the Arrows Task was

much better	better	slightly better	as well	slightly worse	worse	much worse
☐	☐	☐	☐	☐	☐	☐

than my driving performance if I would not have performed the Arrows Task.

I put

much more	some more	slightly more	as much	slightly less	some less	much less
☐	☐	☐	☐	☐	☐	☐

effort in driving than I would have been doing without the Arrows Task.

My driving performance when driving with the system was

much better	better	slightly better	as well	slightly worse	worse	much worse
☐	☐	☐	☐	☐	☐	☐

than my driving performance if I would not have been driving with the system.

I put

much more	some more	slightly more	as much	slightly less	some less	much less
☐	☐	☐	☐	☐	☐	☐

effort in driving than I would have been doing without the system.

Part 3: Workload

The strain that you might experience while performing a task can have different characteristics.

Please read the following descriptions of characteristics carefully and mark how demanding *driving and concurrently performing the Arrows Task* have been for you in terms of these characteristics.

Example:

Mental demands

How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)?

Was the task easy or demanding, simple or complex, exacting or forgiving?

Low High

Physical demands

How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

Low High

Temporal demands

How much time pressure did you feel due to the rate or pace at which the task or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

Low High

Performance

How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

Failure Perfect

Effort

How hard did you have to work (mentally and physically) to accomplish your level of performance?

Low High

Frustration Level

How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during your task?

Low High

Part 4: Attention

Please indicate how much you were aware of the following things during the preceding drive by marking one of the boxes in each line:

	Not aware at all						**Completely aware**
Things happening on the road	☐	☐	☐	☐	☐	☐	☐
The presence of other road users (e.g. oncoming traffic)	☐	☐	☐	☐	☐	☐	☐
The course of the road	☐	☐	☐	☐	☐	☐	☐
Your position in the lane	☐	☐	☐	☐	☐	☐	☐
Road markings (lane markings, safety fences etc.)	☐	☐	☐	☐	☐	☐	☐
Possible dangers	☐	☐	☐	☐	☐	☐	☐
The road surroundings (landscape, trees etc.)	☐	☐	☐	☐	☐	☐	☐
Traffic signs (e.g. speed signs)	☐	☐	☐	☐	☐	☐	☐

The following statements are concerned with your subjective experiences while driving and performing the Arrows Task. Please decide how much they applied to you by marking one of the boxes in each line which best corresponds to your personal feelings.

	Strongly disagree						**Strongly agree**
I found it easy to keep a high level of attention during the whole drive.	☐	☐	☐	☐	☐	☐	☐
I felt distracted from driving due to the Arrows Task.	☐	☐	☐	☐	☐	☐	☐
I paid less attention to the driving task than I would have done in real traffic.	☐	☐	☐	☐	☐	☐	☐
I tried to maintain an acceptable level of driving performance without investing too much effort in it.	☐	☐	☐	☐	☐	☐	☐
I wanted to succeed very much on the Arrows Task.	☐	☐	☐	☐	☐	☐	☐
I didn’t take the Arrows Task too seriously.	☐	☐	☐	☐	☐	☐	☐
It was important for me that the Arrows Task did not have a negative effect on my driving performance.	☐	☐	☐	☐	☐	☐	☐
I could not perform as well as I wanted on the Arrows Task. The driving task required all of my attention.	☐	☐	☐	☐	☐	☐	☐

Part 5: Driver Assistance System

During the preceding drive you drove with an assistance system supporting you in lane keeping. Specifically, this system was designed to prevent you from unintentionally departing from the lane.

Function and Action of the System

In the following we would like to know how you perceived the performance and the capabilities of this system. Please read the questions carefully and mark each scale on the position which best corresponds to your subjective impression of the system's behaviour during the preceding drive.

Example:

Not at all Extremely high

Competence
To what extent did the system prevent you from unintentionally departing from the driving lane (did the system serve its purpose)?

Not at all Extremely high

Comprehensibility
To what extent was the system's behaviour comprehensible?

Not at all Extremely high

Predictability
To what extent could the system's behaviour be predicted from moment to moment?

Not at all Extremely high

Trust
How much did you trust the system?

Not at all Extremely high

Please evaluate the appropriateness of the system's behaviour:

The system intervened: Comments:

- ☐ more often than necessary
- ☐ each time you were going to leave the lane
- ☐ less often than necessary
- ☐ the system never intervened

The system's actions were: Comments:

- ☐ too strong
- ☐ appropriate
- ☐ too weak
- ☐ the system never intervened

The system intervened: Comments:

- ☐ too early
- ☐ in due time
- ☐ too late
- ☐ the system never intervened

Interaction with the System

In this driving simulator study you got to know a system which was designed to assist you in lane keeping.

The following statements reflect how you might have perceived your interaction with this system. Please read the statements carefully and mark how much you agree or disagree with each of them.

	Strongly disagree						**Strongly agree**
I needed a lot of time to learn how the system works.	☐	☐	☐	☐	☐	☐	☐
It was easy to drive with this system.	☐	☐	☐	☐	☐	☐	☐
The system kept me always informed about what it is doing.	☐	☐	☐	☐	☐	☐	☐
I had to permanently monitor the system if it is doing something wrong.	☐	☐	☐	☐	☐	☐	☐
Generally, I felt in control of things.	☐	☐	☐	☐	☐	☐	☐
I could easily identify situations where I had to take over control over the system.	☐	☐	☐	☐	☐	☐	☐
I had no problems to take over control over the system when it was necessary.	☐	☐	☐	☐	☐	☐	☐
There was enough time to react to situations the system could not handle.	☐	☐	☐	☐	☐	☐	☐
I managed to react appropriately to situations the system could not handle.	☐	☐	☐	☐	☐	☐	☐
I felt confident driving with this system.	☐	☐	☐	☐	☐	☐	☐

Assessment of the System and its Effects

In this simulator study you got to know a system which was designed to assist you in lane keeping.

In the following section we want to know how useful you find such a system and how you think it would influence your driving behaviour.

Please mark how much you agree with the following statements:

	Strongly disagree						**Strongly agree**
I enjoyed driving with this system.	☐	☐	☐	☐	☐	☐	☐
The system enhanced my lane keeping performance.	☐	☐	☐	☐	☐	☐	☐
The system enabled me to pay more attention to the Arrows Task than I would have been able to do without the system.	☐	☐	☐	☐	☐	☐	☐
I relied more on the system than I would have done in real traffic.	☐	☐	☐	☐	☐	☐	☐
I think the system would increase safety while driving.	☐	☐	☐	☐	☐	☐	☐
I think the system would make me feel more comfortable while driving.	☐	☐	☐	☐	☐	☐	☐
I think the system would encourage me to pay more attention to things unrelated to driving.	☐	☐	☐	☐	☐	☐	☐
I think the system would help me to focus my attention to the road.	☐	☐	☐	☐	☐	☐	☐
I think driving with this system would make driving more monotonous.	☐	☐	☐	☐	☐	☐	☐
I think the system would encourage me to drive riskier.	☐	☐	☐	☐	☐	☐	☐
I think the system would carry the danger of relying too much on it.	☐	☐	☐	☐	☐	☐	☐

How useful would you find such a system to be for you while driving?

Not useful at all Very useful

☐ ☐ ☐ ☐ ☐

Please explain the reasons for your answer:

How much would you like to have a system like the one you just tested installed in your own car?

Not at all Very much

☐ ☐ ☐ ☐ ☐

How often would you use a system like the one you just tested?

Never Very often

☐ ☐ ☐ ☐ ☐

If you would use the system, in what situations would you use it?

Part 6: Interaction with the Driving Simulator

The steering of the simulator vehicle was

Very much like in a real car ☐ ☐ ☐ ☐ ☐ Very different from a real car

Operating the brake and the gas pedal was

Very much like in a real car ☐ ☐ ☐ ☐ ☐ Very different from a real car

The visibility of the road scene was

Very good ☐ ☐ ☐ ☐ ☐ Very poor

The feeling inside the simulator was

Very much like in a real car ☐ ☐ ☐ ☐ ☐ Very different from a real car

Thank you very much for answering this questionnaire!

We thank you very much for your participation in this study!

APPENDIX B: DEMOGRAPHIC QUESTIONNAIRE

Part 1: Demographic data

Year of birth: __________

Gender: ☐ male ☐ female

Profession, title:

____

In what year did you get your driving licence? __________

Approximately how many kilometres did you drive last year? ____________

Approximately how many kilometres have you been driving since you got your driving licence?

☐ 0 – 10 000 km
☐ 10 001 – 30 000 km
☐ 30 001 – 50 000 km
☐ 50 001 – 100 000 km
☐ 100 001 – 150 000 km
☐ More than 150 000 km

Do you have any experiences with driving in a driving simulator?

☐ yes ☐ no

Part 2: Experience with Driver Assistance Systems

This part is concerned with your experience with so-called *Driver Assistance Systems*.
Driver Assistance Systems are systems inside the car which have the aim to support the driver while driving in order to increase driver comfort and safety.
Such systems may assist the driver in vehicle control, for example by automatically maintaining a safe distance to the vehicle in front. They may also inform the driver about safety-critical situations, for example by warning him or her of a potential collision with another vehicle.

Please indicate below your experiences with *Driver Assistance Systems*.
Please, mark **one** of the alternatives.

Adaptive Cruise Control

Adaptive Cruise Control is a system which automatically holds a set speed and, if necessary, operates the throttle and brakes to maintain a safe distance to the vehicle in front.

How often have you been driving with an *Adaptive Cruise Control* system?

☐ Never
☐ A few times
☐ Often

If you have been driving with an *Adaptive Cruise Control* system already, where was the system installed?

☐ In your own car
☐ In a car owned by another person or in a rental car
☐ In a driving simulator
☐ Have not been driving with an Adaptive Cruise Control system

What is you opinion about *Adaptive Cruise Control*?

Base you answer on the short information above or, if you have been driving with *Adaptive Cruise Control* already, on your experiences with the system.

☐ I would use such a system while driving. I think it is useful.
☐ I might use such a system if it had proven useful to me.
☐ I would not use such a system. I doubt about its usefulness.

Lane Departure Warning

Lane Departure Warning is a system which assists the driver in lane keeping by providing an acoustic or haptic warning signal when he or she unintentionally deviates from the lane.

How often have you been driving with a *Lane Departure Warning* system?

☐ Never
☐ A few times
☐ Often

If you have been driving with a *Lane Departure Warning* system already, where was the system installed?

☐ In your own car
☐ In a car owned by another person or in a rental car
☐ In a driving simulator
☐ Have not been driving with a Lane Departure Warning system

What is you opinion about *Lane Departure Warning*?

Base you answer on the short information above or, if you have been driving with a *Lane Departure Warning* system already, on your experiences with the system.

☐ I would use such a system while driving. I think it is useful.
☐ I might use such a system if it had proven useful to me.
☐ I would not use such a system. I doubt about its usefulness.

Heading Control

Heading Control is a system which intervenes when the driver unintentionally departs from the lane by applying a force on the steering wheel in order to guide the driver back to the centre of the lane.

How often have you been driving with a *Heading Control* system?

☐ Never
☐ A few times
☐ Often

If you have been driving with a *Heading Control* system already, where was the system installed?

☐ In your own car
☐ In a car owned by another person or in a rental car
☐ In a driving simulator
☐ Have not been driving with a Heading Control system

What is you opinion about *Heading Control*?

Base you answer on the short information above or, if you have been driving with a *Heading Control* system already, on your experiences with the system.

☐ I would use such a system while driving. I think it is useful.
☐ I might use such a system if it had proven useful to me.
☐ I would not use such a system. I doubt about its usefulness.

How would you describe your general attitude towards *Driver Assistance Systems*, like those mentioned above?

☐ I would count on these systems while driving. They can surely handle critical driving situations better than I can.
☐ I would rely on such systems only if I knew more exactly how they work and was convinced about their trustworthiness.
☐ I'm sceptical about these systems. I would engage such systems only in situations where I was able to intervene if they fail.
☐ I would not rely on such systems while driving. I prefer to be in control of things on my own.

Part 3: Interest in Technology

To what extent do you agree with the statements below?
Please, **mark the square** in each line which best corresponds to your attitude.

	Don't agree at all						**Agree completely**
I'm keen on testing new technical devices.	☐	☐	☐	☐	☐	☐	☐
I like to have new technical stuff even if I don't really need it.	☐	☐	☐	☐	☐	☐	☐
I'm more interested in *how* technical things work, not if they're functioning perfectly.	☐	☐	☐	☐	☐	☐	☐
I'm *not* attracted by the newest technical things on the market.	☐	☐	☐	☐	☐	☐	☐
I like cars because of their technical equipment.	☐	☐	☐	☐	☐	☐	☐
I do *not* care about *how* technical devices work; the main thing is *that* they work.	☐	☐	☐	☐	☐	☐	☐
I do not care so much about the appearance of my car. The main thing is that it takes me where I want to go.	☐	☐	☐	☐	☐	☐	☐

Part 4: Driving style

How would you describe yourself as a driver?
Each line below contains a pair of adjectives which represent opposite driving styles.
Please **mark the box** in each line which best fits to your own driving style; the closer your mark is to one of the adjectives on each side the more you agree with this description of your usual driving behaviour.

poor	☐	☐	☐	☐	☐	☐	☐	good
slow	☐	☐	☐	☐	☐	☐	☐	fast
anxious	☐	☐	☐	☐	☐	☐	☐	courageous
cautious	☐	☐	☐	☐	☐	☐	☐	risk accepting
calm	☐	☐	☐	☐	☐	☐	☐	emotional
hesitant	☐	☐	☐	☐	☐	☐	☐	determined
nervous	☐	☐	☐	☐	☐	☐	☐	confident
defensive	☐	☐	☐	☐	☐	☐	☐	offensive
observant	☐	☐	☐	☐	☐	☐	☐	unobservant
fore-sighted	☐	☐	☐	☐	☐	☐	☐	impulsive

How would you describe your driving habits?
Please indicate how much you agree with the following statements.

	Don't agree at all						**Agree completely**
I enjoy driving.	☐	☐	☐	☐	☐	☐	☐
I only drive as often as necessary.	☐	☐	☐	☐	☐	☐	☐
If I had a choice I would use the car less often.	☐	☐	☐	☐	☐	☐	☐

Thank you for answering this questionnaire!

APPENDIX C: INSTRUCTIONS FOR DRIVERS IN THE HEADING CONTROL GROUP

1. Study purpose

The aim of this study is to investigate drivers' situation awareness on rural roads under varying conditions. The study will take place in the VTI driving simulator, and is conducted by the Chemnitz University of Technology, Germany, in cooperation with VTI, Linköping, Sweden.

2. Your task

You will drive on a rural road with one lane in each driving direction. The posted speed limit is 70 km/h, and other traffic will appear along the route. It is important that you drive the simulator car and interact with other road users in the same way as you would on a real 70 km/h-road with corresponding character and traffic conditions. This refers to speed choice, manoeuvring, time gaps etc. Avoid experimenting with the simulator and avoid to drive in a way the does not belong to real car driving.

The car has a manual gear shift.

The car is equipped with a Heading Control system – a support system that has not yet been introduced on the market. It helps the driver to keep the car within the driving lane. It applies a counterforce to the steering wheel that guides the car back towards the centre of the driving lane if the car comes too close to the lane boundaries or if it moves into the next lane. The Heading Control system is described in more detail in Appendix A.

While driving, you will have another task to perform. A number of arrows will be shown on a screen mounted on the dashboard. Your task is to determine as fast as possible if there is an arrow pointing upward on the screen. You will respond by pushing a button on the screen. The task is referred to as the "Arrows Task" and is described in Appendix B.

You should try to perform as well as possible on the Arrows Task - without jeopardising traffic safety. You decide on yourself whether it is safe for you to perform the Arrows Task while driving and how you distribute your resources between driving and the Arrows Task. Driving the car in a normal and safe way is the primary task!

3. Procedure

In order to get familiar with the simulator, with the driving conditions and the Heading Control system, you will start with a practise drive of about 10 km in the simulator. After that you will practise the Arrows Task while the simulator car is parked.

After the practise session you will perform two test drives. During these test drives, you will drive with the Heading Control system and additionally perform the Arrows Task when it is presented on the screen in the car. The test drives will take about 15 and 25 min respectively.

Before you start with the practice drive, you will fill in a questionnaire with questions about demographic data, your driving experience and your interest in technology. After each drive in the simulator you will be asked to fill in other questionnaires, which will take about 10 min each time. The questions will deal with how you experienced driving with the Heading Control system and performing the Arrows Task while driving.

The whole study will take about 2 hours.

The procedure is summarised in the scheme below:

Start

	Information and instruction Signing of informed consent	
Questionnaire 1		
	Adjustment of eye tracking system	
1st	Practice drive with the Heading Control system	8 min
Questionnaire 2		
	Training of the Arrows task	
2nd	Test drive with the Heading Control system and the Arrows Task	15 min
Questionnaire 3		
3rd	Test drive with the Heading Control system and the Arrows Task	25 min
Questionnaire 4		

End

2 h

4. Collected data

The collected simulator data and your answers in the questionnaires will be used for scientific purposes only and will be treated strictly confidential. The data will be made anonymous and the results will be presented in a way that makes it impossible to identify how individual participants have answered and performed.

5. Any other issue

The test leader will see and hear you when you drive in the simulator. You can at any time communicate with her via the microphone in the simulator. If you experience any kind of discomfort during the drive, please tell the test leader.

You can at any time interrupt your participation in the study without having to tell the reason.

You will receive 400 SEK for your participation.

To summarise:

- ✓ Drive as you would do on a real 70 km/h-road of corresponding character with the Heading Control system installed in the car.
- ✓ Read the description of the Heading Control system (Appendix A) carefully in order to understand how the driver support system works.
- ✓ When the arrows appear on the screen in the car, answer as fast as you can without jeopardising safety.
- ✓ Read the description of the Arrows Task (Appendix B) carefully in order to be sure how it will be performed.

Questions?

APPENDIX A

This is how the Heading Control system works

The Heading Control (HC) system has the aim to prevent you from unintentionally departing from the driving lane.

As soon as you drive closer to the lane markings, you will feel a counterforce on the steering wheel which guides the car back towards the centre of the lane. If you drive very close to the lane markings the counterforce will be stronger. If you cross the lane markings the HC system will reach its maximum force.

The HC system is always active while driving. However, the system will be switched off when you operate the turn indicator. This tells the system that you intend to leave the lane and the counterforce on the steering wheel will be suppressed. For example, if you want to overtake a vehicle in front of you, you have to use the turn indicator! Then you will feel no counterforce on the steering wheel when you cross the lane marking. During the overtaking manoeuvre the vehicle is acting like a normal car. If you return back to your own lane after the overtaking manoeuvre the HC system will automatically become active again.

The HC system will be automatically switched off when you have crossed the lane markings about half a meter. If you then move back into your own driving lane the system starts working again.

It is principally possible for you to override the HC system if you apply a counterforce to the steering wheel which is greater than the maximum steering force of the HC system.

Please be aware that the HC system is designed to assist you in lane keeping, not to replace you! The counter forces on the steering wheel are not in each case strong enough to keep the vehicle completely in the driving lane without your help!

In order to be able to experience the support by the Heading Control system (the counterforce), you have to keep your hands on the steering wheel!

Any questions about the HC system? – do not hesitate to ask the test leader!

APPENDIX B

Description of the Arrows Task

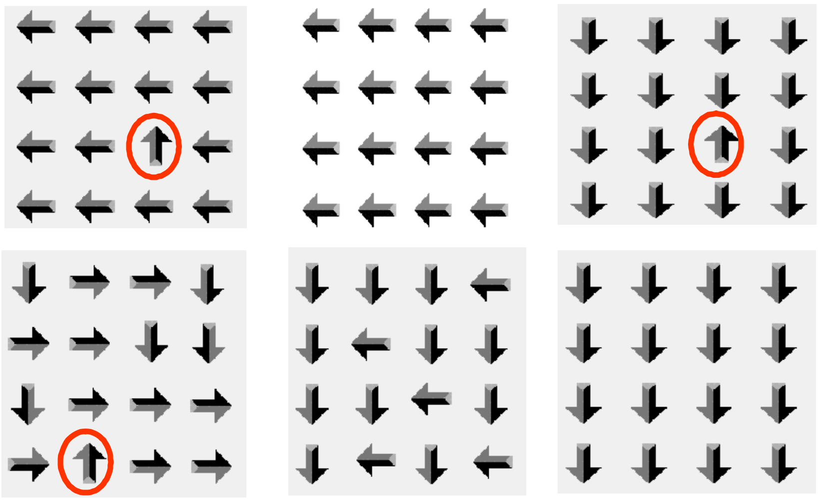
The Arrows Task will be presented occasionally during your drive. Your task will be to determine if a certain arrow appears among a number of arrows presented on a screen mounted on the dashboard.

On the screen 16 arrows will appear. The arrows point in different directions. See the example below. Your task is to determine, as fast as you can, if any of the arrows points upward (circled red in the example below). If there is an arrow pointing upward you have to push the “yes” button in the upper part of the screen. If no arrow points upward you have to push the “no” button in the lower part of the screen.

The Arrows Task lasts for 30 seconds. The start of a new Arrows Task is signalled by a sound.

During the 30 seconds, a new arrows configuration is presented as soon as you have responded by pushing one of the two buttons and your response has been registered. If you do not respond within 5 seconds, a new arrows configuration is presented. After each presentation you will be informed if the response was correct or not. A ”pling” sound means a correct response, while a ”toot” sound means an incorrect response. ”Right” or ”wrong” will also appear on the screen after each response.

Perform the arrows task as well and as fast as possible without jeopardising traffic safety!



Any questions about the Arrows Task? – do not hesitate to ask the test leader!

APPENDIX D: INSTRUCTIONS FOR DRIVERS IN THE LANE DEPARTURE WARNING GROUP

1. Study purpose

The aim of this study is to investigate drivers' situation awareness on rural roads under varying conditions. The study will take place in the VTI driving simulator, and is conducted by the Chemnitz University of Technology, Germany, in cooperation with VTI, Linköping, Sweden.

2. Your task

You will drive on a rural road with one lane in each driving direction. The posted speed limit is 70 km/h, and other traffic will appear along the route. It is important that you drive the simulator car and interact with other road users in the same way as you would on a real 70 km/h-road with corresponding character and traffic conditions. This refers to speed choice, manoeuvring, time gaps etc. Avoid experimenting with the simulator and avoid to drive in a way the does not belong to real car driving.

The car has a manual gear shift.

The car is equipped with a Lane Departure Warning system. It helps the driver to keep the car within the driving lane. It warns the driver by a vibration in the steering wheel when he or she runs the risk of leaving the driving lane. The Lane Departure Warning system is described in more detail in Appendix A.

While driving, you will have another task to perform. A number of arrows will be shown on a screen mounted on the dashboard. Your task is to determine as fast as possible if there is an arrow pointing upward on the screen. You will respond by pushing a button on the screen. The task is referred to as the "Arrows Task" and is described in Appendix B.

You should try to perform as well as possible on the Arrows Task - without jeopardising traffic safety. You decide on yourself whether it is safe for you to perform the Arrows Task while driving and how you distribute your resources between driving and the Arrows Task. Driving the car in a normal and safe way is the primary task!

3. Procedure

In order to get familiar with the simulator, with the driving conditions and the Lane Departure Warning system, you will start with a practise drive of about 10 km in the simulator. After that you will practise the Arrows Task while the simulator car is parked.

After the practise session you will perform two test drives. During these test drives, you will drive with the Lane Departure Warning system and additionally perform the Arrows Task when it is presented on the screen in the car. The test drives will take about 15 and 25 min respectively.

Before you start with the practice drive, you will fill in a questionnaire with questions about demographic data, your driving experience and your interest in technology. After each drive in the simulator you will be asked to fill in other questionnaires, which will take about 10 min each time. The questions deal with how you experienced driving with the Lane Departure Warning system and performing the Arrows Task while driving.

The whole study will take about 2 hours.

The procedure is summarised in the scheme below:

Start

	Information and instruction Signing of informed consent		**2 h**
Questionnaire 1			
	Adjustment of eye tracking system		
1st	Practice drive with the Lane Departure Warning system	8 min	
Questionnaire 2			
	Training of the Arrows task		
2nd	Test drive with the Lane Departure Warning system and the Arrows Task	15 min	
Questionnaire 3			
3rd	Test drive with the Lane Departure Warning system and the Arrows Task	25 min	
Questionnaire 4			

End

4. Collected data

The collected simulator data and your answers in the questionnaires will be used for scientific purposes only and will be treated strictly confidential. The data will be made anonymous and the results will be presented in a way that makes it impossible to identify how individual participants have answered and performed.

5. Any other issue

The test leader will see and hear you when you drive in the simulator. You can at any time communicate with her via the microphone in the simulator. If you experience any kind of discomfort during the drive, please tell the test leader.

You can at any time interrupt your participation in the study without having to tell the reason.

You will receive 400 SEK for your participation.

***To summarise*:**

- ✓ Drive as you would do on a real 70 km/h-road of corresponding character with the Lane Departure Warning system installed in the car.
- ✓ Read the description of the Lane Departure Warning system (Appendix A) carefully in order to understand how the driver support system works.
- ✓ When the arrows appear on the screen in the car, answer as fast as you can without jeopardising safety.
- ✓ Read the description of the Arrows Task (Appendix B) carefully in order to be sure how it will be performed.

Questions?

APPENDIX A

This is how the Lane Departure Warning system works

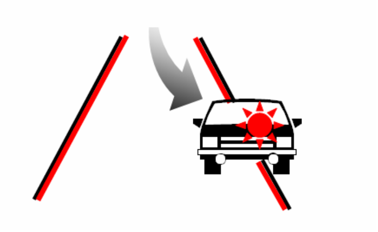
The Lane Departure Warning (LDW) system has the aim to prevent you from unintentionally departing from the driving lane.

As soon as you run the risk of crossing one of the lane markings (side line or centre line) with any part of the car, the LDW system warns you by a vibration in the steering wheel. The vibration lasts for 2 seconds unless you return back to the driving lane within this time. As soon as the car is back within the driving lane, the vibration stops.

When you returned back to the driving lane and you are just about to cross one of the lane markings afterwards again, you will receive a new warning.

The LDW system is always active while driving. However, the system will be switched off when you operate the turn indicator. This tells the system that you intend to leave the lane and the warning will be suppressed. For example, if you want to overtake a vehicle in front of you, you have to use the turn indicator if you want to avoid the warning when crossing the lane marking. During the overtaking manoeuvre the vehicle is acting like a normal car without LDW system. If you return back to your own lane after the overtaking manoeuvre the LDW system will automatically become active again.

In order to be able to experience the support by the LDW system (steering wheel vibrations), you have to keep your hands on the steering wheel!



Any questions about the LDW system? – do not hesitate to ask the test leader!

APPENDIX B

Description of the Arrows Task

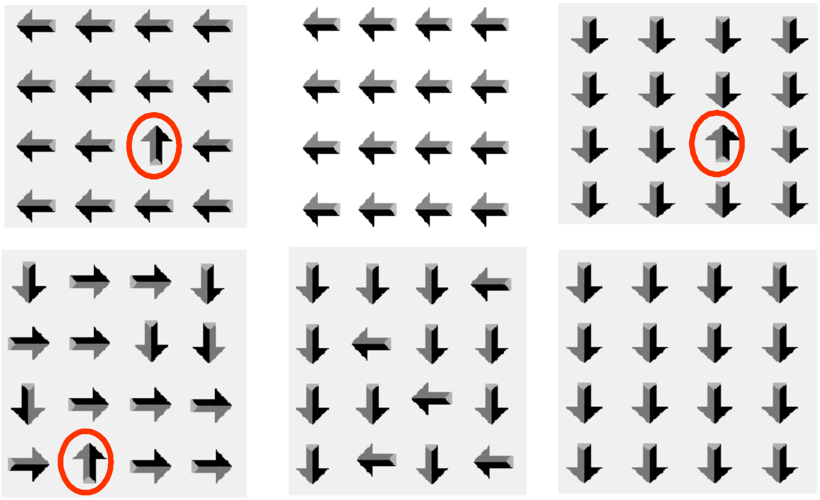
The Arrows Task will be presented occasionally during your drive. Your task will be to determine if a certain arrow appears among a number of arrows presented on a screen mounted on the dashboard.

On the screen 16 arrows will appear. The arrows point in different directions. See the example below. Your task is to determine, as fast as you can, if any of the arrows points upward (circled red in the example below). If there is an arrow pointing upward you have to push the “yes” button in the upper part of the screen. If no arrow points upward you have to push the “no” button in the lower part of the screen.

The Arrows Task lasts for 30 seconds. The start of a new Arrows Task is signalled by a sound.

During the 30 seconds, a new arrows configuration is presented as soon as you have responded by pushing one of the two buttons and your response has been registered. If you do not respond within 5 seconds, a new arrows configuration is presented. After each presentation you will be informed if the response was correct or not. A ”pling” sound means a correct response, while a ”toot” sound means an incorrect response. ”Right” or ”wrong” will also appear on the screen after each response.

Perform the arrows task as well and as fast as possible without jeopardising traffic safety!



Any questions about the Arrows Task? – do not hesitate to ask the test leader!

APPENDIX E: INSTRUCTIONS FOR DRIVERS IN THE CONTROL GROUP

1. Study purpose

The aim of this study is to investigate drivers' situation awareness on rural roads under varying conditions. The study will take place in the VTI driving simulator, and is conducted by the Chemnitz University of Technology, Germany, in cooperation with VTI, Linköping, Sweden.

2. Your task

You will drive on a rural road with one lane in each driving direction. The posted speed limit is 70 km/h, and other traffic will appear along the route. It is important that you drive the simulator car and interact with other road users in the same way as you would on a real 70 km/h-road with corresponding character and traffic conditions. This refers to speed choice, manoeuvring, time gaps etc. Avoid experimenting with the simulator and avoid to drive in a way the does not belong to real car driving.

The car has a manual gear shift.

While driving, you will have another task to perform. A number of arrows will be shown on a screen mounted on the dashboard. Your task is to determine as fast as possible if there is an arrow pointing upward on the screen. You will respond by pushing a button on the screen. The task is referred to as the "Arrows Task" and is described in Appendix A.

You should try to perform as well as possible on the Arrows Task - without jeopardising traffic safety. You decide on yourself whether it is safe for you to perform the Arrows Task while driving and how you distribute your resources between driving and the Arrows Task. Driving the car in a normal and safe way is the primary task!

3. Procedure

In order to get familiar with the simulator and with the driving conditions, you will start with a practise drive of about 10 km in the simulator. After that you will practise the Arrows Task while the simulator car is parked.

After the practise session you will perform two test drives. During these test drives, you will additionally perform the Arrows Task while driving when it is presented on the screen in the car. The test drives will take about 15 and 25 min respectively.

Before you start with the practice drive, you will fill in a questionnaire with questions about demographic data, your driving experience and your interest in technology. After each drive in the simulator you will be asked to fill in other questionnaires, which will take about 10 min each time. The questions deal with how you experienced driving and performing the Arrows Task while driving.

The whole study will take about 2 hours.

The procedure is summarised in the scheme below:

Start

	Information and instruction Signing of informed consent	
Questionnaire 1		
	Adjustment of eye tracking system	
1st	Practice drive	8 min
Questionnaire 2		
	Training of the Arrows task	
2nd	Test drive with the Arrows Task	15 min
Questionnaire 3		
3rd	Test drive with the Arrows Task	25 min
Questionnaire 4		

End

2 h

4. Collected data

The collected simulator data and your answers in the questionnaires will be used for scientific purposes only and will be treated strictly confidential. The data will be made anonymous and the results will be presented in a way that makes it impossible to identify how individual participants have answered and performed.

5. Any other issue

The test leader will see and hear you when you drive in the simulator. You can at any time communicate with her via the microphone in the simulator. If you experience any kind of discomfort during the drive, please tell the test leader.

You can at any time interrupt your participation in the study without having to tell the reason.

You will receive 400 SEK for your participation.

To summarise:

- ✓ Drive as you would do on a real 70 km/h-road of corresponding character.
- ✓ When the arrows appear on the screen in the car, answer as fast as you can without jeopardising safety.
- ✓ Read the description of the Arrows Task (Appendix A) carefully in order to be sure how it will be performed.

Questions?

APPENDIX A

Description of the Arrows Task

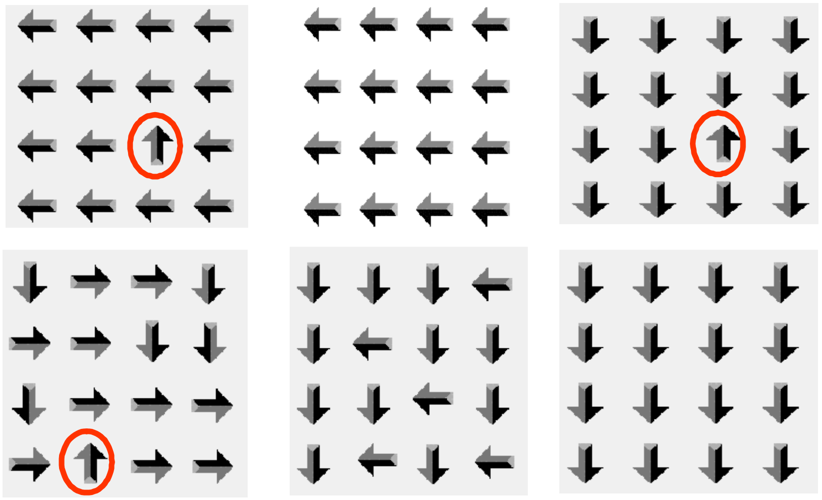
The Arrows Task will be presented occasionally during your drive. Your task will be to determine if a certain arrow appears among a number of arrows presented on a screen mounted on the dashboard.

On the screen 16 arrows will appear. The arrows point in different directions. See the example below. Your task is to determine, as fast as you can, if any of the arrows points upward (circled red in the example below). If there is an arrow pointing upward you have to push the “yes” button in the upper part of the screen. If no arrow points upward you have to push the “no” button in the lower part of the screen.

The Arrows Task lasts for 30 seconds. The start of a new Arrows Task is signalled by a sound.

During the 30 seconds, a new arrows configuration is presented as soon as you have responded by pushing one of the two buttons and your response has been registered. If you do not respond within 5 seconds, a new arrows configuration is presented. After each presentation you will be informed if the response was correct or not. A ”pling” sound means a correct response, while a ”toot” sound means an incorrect response. ”Right” or ”wrong” will also appear on the screen after each response.

Perform the arrows task as well and as fast as possible without jeopardising traffic safety!



Any questions about the Arrows Task? – do not hesitate to ask the test leader!

APPENDIX F: LANE KEEPING PERFORMANCE DATA

Table F1. Drivers' mean standard deviation of lateral position [m] in driving periods without and with concurrent Arrows Task performance in Session 1 and Session 2 as a function of the level of lane keeping assistance.

		Session 1				Session 2			
		No Arrows		Arrows		No Arrows		Arrows	
LoA	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
HC	15	0.11	0.02	0.15	0.03	0.12	0.02	0.14	0.03
LDW	14	0.11	0.02	0.16	0.02	0.13	0.02	0.16	0.02
NA	15	0.12	0.02	0.17	0.02	0.14	0.02	0.16	0.02

Note. LoA = Level of lane keeping assistance.

Table F2. Mean duration of maximum HC steering torque [s] in about 28 s-long driving periods without and with concurrent Arrows Task performance in Session 1 and Session 2 as a function of the level of lane keeping assistance.

		Session 1				Session 2			
		No Arrows		Arrows		No Arrows		Arrows	
LoA	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
HC	15	0.06	0.16	0.22	0.29	0.02	0.05	0.08	0.14
LDW	15	0.15	0.35	0.56	0.62	0.16	0.23	0.44	0.38
NA	15	0.17	0.34	0.89	0.57	0.38	0.55	0.56	0.77

Note. LoA = Level of lane keeping assistance.

APPENDIX G: RELIANCE DATA

Table G1. Mean proportion of display glance time in normal driving periods in Session 1 and Session 2 and in critical driving situations (Events) in Session 2 as a function of the level of lane keeping assistance, separately for the two subgroups of HC drivers that were found to differ significantly in their mean single display glance durations.

		Session 1		Session 2		Events	
LoA	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
HC long	5	.72	.07	.64	.10	.60	.08
HC short	10	.66	.08	.60	.11	.55	.10
LDW	12	.70	.09	.65	.09	.60	.10
NA	14	.70	.09	.65	.10	.61	.09

Note. LoA = Level of lane keeping assistance.

APPENDIX H: SPEED DATA

Table H1. Drivers' mean speed [km/h] in driving periods with concurrent Arrows Task performance in Session 1 and Session 2 as a function of the level of lane keeping assistance, separately for the two subgroups of HC drivers that were found to differ significantly in their reliance on the HC system.

		Session 1		Session 2	
LoA	*n*	*M*	*SD*	*M*	*SD*
HC long	5	73.32	4.38	76.27	6.78
HC short	9	73.34	4.36	72.29	4.06
LDW	15	74.17	5.36	75.05	5.45
NA	14	74.12	5.37	75.18	5.49

Note. LoA = Level of lane keeping assistance.

APPENDIX I: MENTAL EFFORT REGULATION DATA

Table I1. Drivers' mean response times [s] to the Arrows Task displays during normal driving periods in Session 1 and Session 2 and during critical driving situations (Events) as a function of the level of lane keeping assistance, separately for the two subgroups of HC drivers that were found to differ significantly in their reliance on the HC system.

		Session 1		Session 2		Events	
LoA	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
HC long	5	3.42	0.20	3.38	0.35	3.74	0.54
HC short	10	2.87	0.77	2.84	0.81	3.24	0.89
LDW	15	2.95	0.64	2.92	0.66	3.40	0.69
NA	15	2.95	0.55	2.91	0.50	3.35	0.59

Note. LoA = Level of lane keeping assistance.

Table I2. Drivers' mean percentage [%] of correct responses to the Arrows Task during normal driving periods in Session 1 and Session 2 and during critical driving situations (Events) as a function of the level of lane keeping assistance.

		Session 1		Session 2		Events	
LoA	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
HC	15	87.98	8.74	86.71	9.89	68.02	20.77
LDW	15	89.59	8.23	89.14	11.07	71.37	19.39
NA	15	91.37	7.84	92.80	8.38	74.49	12.22

Note. LoA = Level of lane keeping assistance.

Table I3. Drivers' mean percentage [%] of false responses to the Arrows Task during normal driving periods in Session 1 and Session 2 and during critical driving situations (Events) as a function of the level of lane keeping assistance.

		Session 1		Session 2		Events	
LoA	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
HC	15	10.06	5.22	11.28	8.42	22.50	14.21
LDW	15	9.42	6.93	9.77	9.35	24.81	16.88
NA	15	7.80	6.26	6.21	5.87	20.29	9.25

Note. LoA = Level of lane keeping assistance.

Table I4. Drivers' mean percentage [%] of missing responses to the Arrows Task during normal driving periods in Session 1 and Session 2 and during critical driving situations (Events) as a function of the level of lane keeping assistance.

		Session 1		Session 2		Events	
LoA	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
HC	15	1.58	3.46	1.18	2.18	8.75	10.34
LDW	15	0.32	0.73	0.51	1.40	3.14	4.62
NA	15	0.70	2.04	0.87	2.32	3.92	4.45

Note. LoA = Level of lane keeping assistance.

Table I5. Mean response accuracy index [0 – 1] as a combined measure of drivers' percentages of correct, false, and missing responses to the Arrows Task during normal driving periods in Session 1 and 2 and during critical driving situations (Events) as a function of the level of lane keeping assistance.

		Session 1		Session 2		Events	
LoA	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
HC	15	.93	.05	.93	.05	.82	.13
LDW	15	.94	.04	.94	.06	.85	.10
NA	15	.95	.04	.96	.05	.86	.07

Note. LoA = Level of lane keeping assistance.

Table I6. Drivers' ratings of overall mental workload on the Rating Scale Mental Effort (RSME) for the practice session and for Session 1 and Session 2 as a function of the level of lane keeping assistance, separately for the two subgroups of HC drivers that were found to differ significantly in their reliance on the HC system.

		Practice		Session 1		Session 2	
LoA	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
HC long	5	30.40	5.77	61.40	14.08	87.20	16.90
HC short	10	26.60	14.65	57.40	30.24	72.70	24.30
LDW	15	31.73	15.03	59.07	27.51	75.60	32.02
NA	15	27.13	18.07	45.73	18.85	61.87	22.19

Note. LoA = Level of lane keeping assistance.

Table I7. Drivers' ratings of overall mental workload [0 – 100] as measured by the Raw Task Load Index (RTLX) for Session 1 and Session 2 as a function of the level of lane keeping assistance, separately for the two subgroups of HC drivers that were found to differ significantly in their reliance on the HC system.

		Session 1		Session 2	
LoA	*n*	*M*	*SD*	*M*	*SD*
HC long	5	55.90	10.01	65.63	9.24
HC short	10	44.04	15.17	53.75	14.17
LDW	15	47.02	9.60	60.19	12.56
NA	15	45.09	11.34	49.41	14.10

Note. LoA = Level of lane keeping assistance.

Table I8. Drivers' workload ratings [0 – 100] on six dimensions of mental workload as measured by the NASA Task Load Index (NASA-TLX) for Session 1 and Session 2 as a function of the level of lane keeping assistance, separately for the two subgroups of HC drivers that were found to differ significantly in their reliance on the HC system.

		Session 1		Session 2	
LoA	*n*	*M*	*SD*	*M*	*SD*
		Mental demands			
HC long	5	61.00	17.78	76.40	10.88
HC short	10	49.60	20.32	66.40	15.38
LDW	15	52.93	15.21	68.37	16.40
NA	15	52.73	17.71	56.13	13.30
		Physical demands			
HC long	5	24.20	12.68	42.60	17.87
HC short	10	26.40	19.43	43.00	25.91
LDW	15	25.00	17.98	48.23	19.79
NA	15	18.80	14.51	32.40	19.22
		Temporal demands			
HC long	5	64.40	15.53	78.20	13.85
HC short	10	42.15	29.18	51.30	28.71
LDW	15	51.67	22.13	64.27	22.42
NA	15	48.67	20.56	55.00	22.19

Continuation of Table I8.

		Session 1		Session 2	
LoA	*n*	*M*	*SD*	*M*	*SD*
		Performance			
HC long	5	54.40	19.73	50.40	11.93
HC short	10	68.15	22.37	57.30	20.53
LDW	15	71.40	19.10	62.07	15.95
NA	15	66.27	18.45	61.67	23.54
		Effort			
HC long	5	65.20	8.70	77.60	8.56
HC short	10	51.55	26.50	63.40	20.08
LDW	15	58.20	18.81	72.93	14.30
NA	15	53.33	19.19	53.73	21.65
		Frustration			
HC long	5	66.20	18.32	68.60	17.98
HC short	10	26.40	21.36	41.10	22.16
LDW	15	22.93	22.69	45.27	28.09
NA	15	30.73	23.20	37.53	27.21

Note. LoA = Level of lane keeping assistance.

Table I9. Analysis of variance for the NASA-TLX workload dimension "Mental demands".

Source	*df*	*F*	η^2_G	*p*
		Between subjects		
Level of lane keeping assistance (LoA)	3	1.36	.07	.27
S within-group error	41	(408.17)		
		Within subjects		
Session (Se)	1	26.13**	.12	.00
Se x LoA	3	2.28	.04	.09
Se x S within-group error	41	(115.00)		

Note. Values enclosed in parentheses represent mean square errors. S = subjects.
$^{**}p < .01$.

Table I10. Analysis of variance for the NASA-TLX workload dimension "Physical demands".

Source	*df*	*F*	η^2_G	*p*
		Between subjects		
Level of lane keeping assistance (LoA)	3	1.13	.06	.35
S within-group error	41	(599.82)		
		Within subjects		
Session (Se)	1	49.92**	.17	.00
Se x LoA	3	1.01	.01	.40
Se x S within-group error	41	(119.28)		

Note. Values enclosed in parentheses represent mean square errors. S = subjects.
$^{**}p < .01$.

Table I11. Analysis of variance for the NASA-TLX workload dimension "Temporal demands".

Source	*df*	*F*	η^2_G	*p*
Between subjects				
Level of lane keeping assistance (LoA)	3	1.99	.10	.13
S within-group error	41	(767.26)		
Within subjects				
Session (Se)	1	6.88*	.04	.01
Se x LoA	3	0.22	.00	.88
Se x S within-group error	41	(294.29)		

Note. Values enclosed in parentheses represent mean square errors. S = subjects.
*$p < .05$.

Table I12. Analysis of variance for the NASA-TLX workload dimension "Performance".

Source	*df*	*F*	η^2_G	*p*
Between subjects				
Level of lane keeping assistance (LoA)	3	0.96	.05	.42
S within-group error	41	(543.92)		
Within subjects				
Session (Se)	1	4.21*	.03	.05
Se x LoA	3	0.26	.01	.86
Se x S within-group error	41	(227.30)		

Note. Values enclosed in parentheses represent mean square errors. S = subjects.
*$p < .05$.

Table I13. Analysis of variance for the NASA-TLX workload dimension "Effort".

Source	*df*	*F*	η^2_G	*p*
		Between subjects		
Level of lane keeping assistance (LoA)	3	2.02	.12	.13
S within-group error	41	(594.73)		
		Within subjects		
Session (Se)	1	12.88**	.06	.00
Se x LoA	3	2.10	.03	.12
Se x S within-group error	41	(138.94)		

Note. Values enclosed in parentheses represent mean square errors. S = subjects.
***p* < .01.

Table I14. Analysis of variance for the NASA-TLX workload dimension "Frustration".

Source	*df*	*F*	η^2_G	*p*
		Between subjects		
Level of lane keeping assistance (LoA)	3	3.72*	.17	.02
S within-group error	41	(887.18)		
		Within subjects		
Session (Se)	1	9.24**	.05	.00
Se x LoA	3	1.54	.03	.22
Se x S within-group error	41	(266.82)		

Note. Values enclosed in parentheses represent mean square errors. S = subjects.
**p* < .05. ***p* < .01.

Table I15. Mean percentage of overall effort [%] that drivers invested in the driving task in Session 1 and Session 2 as a function of the level of lane keeping assistance, separately for the two subgroups of HC drivers that were found to differ significantly in their reliance on the HC system.

		Session 1		Session 2	
LoA	*n*	*M*	*SD*	*M*	*SD*
HC long	5	47.60	14.93	62.80	10.69
HC short	10	51.30	25.82	74.70	12.50
LDW	15	42.47	19.84	67.47	18.43
NA	15	53.60	18.54	64.93	16.39

Note. LoA = Level of lane keeping assistance.

Table I16. Drivers' mean ratings on the item "It was important for me that the Arrows Task did not have a negative effect on my driving performance" [1 – strongly disagree; 7 – strongly agree] in Session 1 and Session 2 as a function of the level of lane keeping assistance.

		Session 1		Session 2	
LoA	*n*	*M*	*SD*	*M*	*SD*
HC	15	4.80	1.47	4.73	1.49
LDW	15	5.33	1.23	5.60	0.99
NA	15	5.67	0.90	6.20	1.01

Note. LoA = Level of lane keeping assistance.

Table I17. Drivers' mean ratings on the item "I wanted to succeed very much on the Arrows Task" [1 – strongly disagree; 7 – strongly agree] in Session 1 and Session 2 as a function of the level of lane keeping assistance, separately for the two subgroups of HC drivers that were found to differ significantly in their reliance on the HC system.

		Session 1		Session 2	
LoA	*n*	*M*	*SD*	*M*	*SD*
HC long	5	6.20	0.84	5.40	0.89
HC short	10	5.50	1.18	4.50	1.65
LDW	15	6.00	0.93	4.67	1.95
NA	15	5.87	1.13	5.13	1.46

Note. LoA = Level of lane keeping assistance.

APPENDIX J: BASIS OF TRUST DATA

Table J1. Drivers' evaluation of the HC system's and the LDW system's competence, comprehensibility and predictability [0 – not at all, 100 – extremely high] for Session 1 and Session 2, separately for the two subgroups of HC drivers that were found to differ substantially in their trust in and in their reliance on the HC system.

		Session 1		Session 2	
LoA	*n*	*M*	*SD*	*M*	*SD*
		Competence[a]			
HC long	5	83.80	14.24	68.80	19.83
HC short	10	59.30	32.72	62.10	26.57
LDW	15	52.33	28.41	59.20	27.22
		Comprehensibility[b]			
HC long	5	81.60	10.57	74.80	17.99
HC short	10	91.00	11.47	77.10	15.16
LDW	15	87.33	12.98	84.87	14.41
		Predictability[c]			
HC long	5	72.80	10.21	67.40	13.63
HC short	10	66.50	27.20	70.60	18.48
LDW	15	66.53	31.53	68.50	25.98

Note. LoA = Level of lane keeping assistance.

[a] "To what extent did the system prevent you from unintentionally departing from the driving lane (did the system serve its purpose)?". [b] "To what extent was the system's behaviour comprehensible?". [c] "To what extent could the system's behaviour be predicted from one moment to the next?".

APPENDIX K: ENERGETIC AROUSAL DATA

Table K1. Drivers' ratings on the Stanford Sleepiness Scale (1 – 7) for the practice session (10-min drive without secondary task), for Session 1 (15-min drive with secondary task) and for Session 2 (25-min drive with secondary task and occurrence of critical driving situations) as a function of the level of lane keeping assistance, separately for the two subgroups of HC drivers that were found to differ significantly in their reliance on the HC system.

		Practice		Session 1		Session 2	
LoA	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
HC long	5	2.60	0.55	2.80	0.45	2.40	0.55
HC short	10	2.60	0.70	2.40	0.70	2.30	0.68
LDW	14	2.21	0.89	2.07	0.88	2.13	0.83
NA	14	1.36	0.63	1.64	0.63	1.86	0.77

Note. LoA = Level of lane keeping assistance.

APPENDIX L: SITUATION AWARENESS DATA

Table L1. Drivers' awareness ratings [1 – not aware at all, 7 – completely aware] for different types of driving-related information in Session 1 and Session 2 as a function of the level of lane keeping assistance.

		Session 1		Session 2	
LoA	*n*	*M*	*SD*	*M*	*SD*
		Things happening on the road			
HC	15	4.87	1.46	5.47	1.25
LDW	15	5.87	0.99	6.07	0.46
NA	15	5.93	0.70	6.20	0.68
		Presence of other road users			
HC	15	5.13	1.55	5.67	1.23
LDW	15	6.13	1.06	6.27	0.70
NA	15	6.07	1.10	6.33	0.62
		Course of the road			
HC	15	4.73	1.79	5.33	1.34
LDW	15	5.71	1.14	5.43	1.45
NA	15	5.80	1.15	5.67	1.23
		Position in the lane			
HC	15	4.33	1.50	5.07	1.03
LDW	15	5.40	0.63	5.20	0.78
NA	15	5.13	0.99	5.47	0.74

Continuation of Table L1.

		Session 1		Session 2	
LoA	*n*	*M*	*SD*	*M*	*SD*
Road markings					
HC	15	4.67	1.50	5.13	1.19
LDW	15	5.13	0.83	5.07	0.96
NA	15	5.47	1.19	5.67	0.90

Note. LoA = Level of lane keeping assistance.