Gaps in the control of automated vehicles on roads

Calvert, Simeon; Mecacci, Giulio; van Arem, Bart; Santoni De Sio, Filippo; Heikoop, Daniël; Hagenzieker, Marjan

DOI
10.1109/MITS.2019.2926278

Publication date
2020

Document Version
Final published version

Published in
IEEE Intelligent Transportation Systems Magazine

Citation (APA)
Calvert, S., Mecacci, G., van Arem, B., Santoni De Sio, F., Heikoop, D., & Hagenzieker, M. (2020). Gaps in the control of automated vehicles on roads. *IEEE Intelligent Transportation Systems Magazine*, *13 (2021)*(4), 146-153. https://doi.org/10.1109/MITS.2019.2926278

Gaps in the Control of Automated Vehicles on Roads

Simeon C. Calvert*, Bart van Arem, Daniël D. Heikoop, and Marjan Hagenzieker
Department of Transport & Planning, Delft University of Technology, Delft, The Netherlands.
E-mail: s.c.calvert@tudelft.nl

Giulio Mecacci and Filippo Santoni de Sio
Section of Ethics and Philosophy of Technology, Delft University of Technology, Delft, The Netherlands.

xxxxxxx

Abstract—Increased on-road testing and market availability of partially automated vehicles (AV) offers researchers and developers the opportunity to evaluate the AV's performance. The occurrence of new types of accidents involving AV's has sparked questions in regard to who is actually in control over and responsible for AV control. In this contribution, we suggest a potential discrepancy in AV control with the review of recently documented accidents involving AV's. The identification of a gap in control is performed using a recently formulated moral philosophical framework of Meaningful Human Control (MHC). This shows a discrepancy between the attribution of responsibility and the ability of a human to fulfil the role assigned to them. While a gap in control is not evident from the viewpoint of operational control, it requires the more intricate concept of MHC to expose it. Recommendations are further made that AV developers and vehicle approval authorities should consider control from a MHC perspective to avoid future gaps in control with the resulting consequences.

Digital Object Identifier 10.1109/MITS.2019.2926278
Date of current version: 31 January 2020

*Corresponding author

1939-1390/20©2020IEEE

I. Introduction

On-road testing and increased automated functionality in production road vehicles has increased steeply in recent years. It is estimated that many hundreds of million miles have now been driven in vehicles that have SAE level 2 capability or higher (both longitudinal and lateral automation, but with the driver monitoring) [4], [5]. However, accidents involving automated vehicles (AV's) have also been occurring and have attracted increased media attention. It is not surprising that accidents occur. However the causes behind the accidents do give insight into the performance of the current crop of AV's on roads and there may be cause for concern. While responsibility is quickly attributed by various parties, there may to be a deeper underlying problem in regard to AV-control. In this article, we aim to address the aspect of control over an automated vehicle and show that current driver-vehicle setups may contain a critical gap in their control chain.

Quotations of the number of accidents involving AV's vary extensively depending on the source, but can be found to lie in the region of one accident per 42017 miles [4], while it remains inconclusive if AV's are safer than conventional vehicles due to low and non-representative conditions [8]. The vast majority of the accidents are at very low speeds (<10 mph) with minimal to no structural damage, never mind human injury or death [4]. The first reported deaths involving an AV on public roads have also occurred. Three well publicized incidents have been the Tesla-trailer collision in May 2016 in Florida, the Uber Volvo collision with a pedestrian in March 2018 in Arizona and the Tesla collision with a parked police car in May 2018 in California. [1], [3], [6], [7]. Characteristics of the first two incidents are given in Table I based on official reports. The official investigation report for the third accident is pending. For this reason, use is made of official police statements gathered by media. In each of these incidents, a similar explanation emerged: the vehicle was not able to fulfil a designated task and the driver did or could not react to mitigate the impending incident. Without further analysis, we can already clearly state that a discord existed between the driver and the Automated Driving Control System (ADCS)[1]. The three accidents are analyzed later in the article and are used for a proxy of current AV systems in practice.

In all cases, we are considering low level automation here. For low level automated vehicles, SAE [9] describe SAE L1 AV's as vehicles that have automated lateral or longitudinal control within their Operational Design Domain (ODD), and SAE L2 vehicles as those that have both. The dynamic Driving Task (DDT) is not performed by the ADCS, but by the driver, as the ADCS only performs part of the DDT [9]. Object and Event Detection and Response (OEDR), as defined by SAE, are the responsibility of the driver, even if the ADCS performs some of the tasks. Partially automated SAE L3 AV's perform the whole DDT, but only within their ODD. The driver performs a fall-back role and must be receptive to intervene in a timely fashion [9].

With increasing amounts of behavioral and psychological research focusing on the role of drivers in AV's,

[1]We define Automated Driving Control System (ADCS) as the complete setup of an automated vehicles' control algorithms, software and related sensory hardware and ability.

Table I. Accident characteristics of three selected AV accidents.

Incident	Tesla (T)-Tractor Trailer (TT) Collision-Florida May 2016 [1], [2]	Uber Volvo (UV)-Pedestrian (P) Collision-Arizona March 2018 [3]	Tesla (T)-Police Car (PC) Collision-California May 2018 [6], [7]
Road type	Rural highway	Urban street, 2 lanes	Urban street
Accident type	Side on collision T on TT at uncontrolled interaction	UV collides side with crossing P	T collides with road-side parked PC
Mortality/Injury	Driver of T dies	P dies	T driver minor injuries
ADCS system in operation	Tesla autopilot mode (SAE L2)	Uber 'developmental self-driving system'	Tesla autopilot mode (SAE L2)
Stated role of driver	Continued and full attention of driver to monitor… and be prepared to take action to avoid crashes.	Attentive operator to intervene if system fails to perform appropriately during testing.	Continued and full attention of driver to monitor… and be prepared to take action to avoid crashes.
ADCS performance	T 'Automatic Emergency Braking' did not warn or perform braking action. System was found to be working as designed.	UV system detected P 6s before collision. System determined braking action required at 1.3s prior to collision, but did actuate. Emergency braking maneuvers were disabled. System was operating normally without faults.	Unknown
Driver state	Unknown	Distracted	Unknown
Driver action	No action detected	Braking action less than 1s before collision	Unknown
Probable cause of accident	T perception sensors did not detect TT	Disabled emergency braking system and inattentive driver	Possibly vehicle following unclear road marking (not confirmed!)

increasingly more evidence is appearing that suggests that drivers are not suitably fit to perform the tasks that are demanded of them in AV's [10]-[12]. In general, this follows the line of reasoning that drivers whose tasks are reduced to only monitoring are subject to a reduction in situational awareness (SA) [13]-[16] as they experience a lack of task demand and intensity, which has been shown to lead to a short-term degradation in Task capability [16]-[18]. SA describes the processes of attention, perception, and decision making in regard to a person's mental model of their current situation [19], [20]. Reduced SA can lead to inattention and even distraction during their monitoring tasks [13], [21].

This is not purely by choice, but is inherent to the way human cognitive processes work. As a consequence, the quality of the performance of monitoring decreases [16], [22]-[25], which leads to longer reaction times and even incoherent reactions to stimuli [10], [26]. And if an ADCS makes a request for a driver to (immediately) retake operational control of a vehicle, there is ample evidence that a driver shows a significant inability to do this in a timely and correct fashion in emergency situations [12], [26]-[29]. In regard to control, and therefore responsibility, often too much is expected from drivers to perform tasks that they cannot reasonably be expected to perform [10]-[12], [30].

Automated vehicle manufacturers have repeatedly stated that their current vehicles are not able to drive fully automated or autonomously [1], [31] and require drivers to remain vigilant and resume operational control if required. However, it is clear that current AV's cannot be deemed to always be capable of performing driving tasks, as also demonstrated from recent accidents. However, with questions regarding a driver's ability to sufficiently fulfil their DDT while driving with low automation, and also regarding the suitability and clarity of the applied distribution of tasks and responsibility between DDT, OEDR and 'Fall back' in practice, we again arrive at our thesis that a gap in control exists in the current design and operation of AV's. We aim to address the concept of control in the following section. Thereafter, we argue that the use of the concept of Meaningful Human Control (MHC) is required to identify and bridge the gaps in control, and we discuss consequences thereof in the discussions and recommendations.

II. Automated Vehicle Control

A. Levels of Control

A classical description of skills and control was coined by Michon [32], which distinguished between strategical, tactical and operational levels of control. The strategical level defines the general planning of a trip, such as route choice, mode choice, etc., the tactical level involves driver maneuvers, while the operational level is the physical action of movement at any one time. Although control can be discussed on all three levels, we are going to focus here on operational control, as this is the level at which actions are performed and that must be considered most critical, e.g. for safety and alike. If failures occur on an operational level, then incorrect vehicle movements are the consequence, which directly create unsafe situations and can lead to accidents. Operational control is performed over the Dynamic Driving Task (DDT) as defined by SAE [9] and aligns with their description of the DDT.

Under normal driving conditions in a conventional vehicle without any form of automation, as well as strategic and tactical control, a driver should be in operational control at all times, i.e. they should actively perform all sorts of driving tasks. Consequentially, the driver can be considered to be responsible for the behavior of the vehicle as they are required to be in control. The other extreme is in fully automated vehicles, in which a human driver is not required at any time and the vehicle performs all DDT including monitoring of the environment (e.g. OEDR). In this case, which does not yet exist on roads or even in most field trials, the ADCS is in complete control. In both these cases, the vehicle is designed such that control is carried out as envisaged. Intermediate levels of automation that a) perform some driving tasks (such as ACC) and leave the rest to a driver, or b) perform most or all tasks, but require a driver to monitor and if required intervene, remove some or all driving tasks from a driver, while still demand the driver to remain engaged [9]. The driver is left with only an observatory control task, which, as stated, leads to a loss in task capability.

B. Meaningful Human Control

Recent developments in philosophy of technology have led to a new concept of control: Meaningful Human Control (MHC), which has been backed as a vital and necessary concept for vehicle automation going forward [33], [34]. MHC entails the extent to which humans can maintain control over an (automated) system, even when not actively performing driving tasks, for example by means of system design. This notion was defined by Santoni de Sio and Van den Hoven [34] in the context of the political debate on autonomous weapon systems, to depart from an idea of direct operational control of an agent over an intelligent system, towards control mechanisms that originate from human reasons to act. The application of the concept of MHC in vehicle automation is logical as humans must maintain generic control over such an ADCS that is there to aid mobility, but also has the potential to cause undesirable, unsafe or even dangerous situations [35]. The concept of MHC relies on two formal conditions called *tracking* and *tracing*. The *tracking condition* considers the responsiveness of a system to act according to human reasons. This condition denotes any factor that can motivate and explain human behavior, such as intentions and plans. For example, the intention of a driver might be to get home as soon as possible and therefore decide to ignore a stop sign (note, we are not evaluating intentions, just recognizing them).We would hope than an automated system would not follow such an intention and stop at the sign. The *tracing condition* demands the possibility to identify one or more human agents (e.g. ADCS designers, drivers, etc.) in the system's design and operation, who are able to: (i) appreciate the capabilities of the system and (ii) understand their own role as targets of potential moral consequences for the system's behavior. This could be the driver, but does not have to be. MHC defines conditions for control that do not depend on whether a particular agent is performing specific tasks. Rather, those conditions regard certain capacities of the system as a whole. In such a way, it is clear that operational control by a qualified driver can lead to the system being under MHC, not just because a driver engages in driving tasks, but because the system satisfies the two fundamental conditions. The areas where control may be perceived to lie in theory are given in Table II. The extent to which this is really the case is tested in the following Section.

Table II. Agent in control of automated vehicle.

Level of Automation	Operational Control	Monitoring	Agent With Potential MHC Over System*
None (SAE L0)	Driver	Driver	Driver
Low (SAE L1-3)	ADCS (+ driver)	Driver	Driver & Designer of ADCS
High (SAE L4-5)	ADCS	ADCS	Designer of ADCS

*note that this list is illustrative and as a consequence also restrictive, as in practice MHC also responds to other relevant reasons, including societal reasons as embedded in infrastructures, road signs, regulations etc.

III. Gaps in Control Over Automated Vehicles

A. Analysis of Operational Control

As a proxy for partially automated vehicles in practice, let us again consider the three serious accidents that we earlier referred to and are characterized in Table I. In each of the cases, the vehicle could be considered a low level automated vehicle that is capable of performing all driving tasks within the specified driving conditions on the applicable road, while requiring the driver to maintain vigilance and monitor the environment and be prepared to intervene if required by either the circumstances or on request by the vehicle. The vehicle manufacturers in each case are very clear on the ODD of their vehicles. The Tesla vehicles from the first and third "require the continual and full attention of the driver to monitor the traffic environment" [1], [2], as "many unforeseen circumstances can impair the operation … and as a result may not steer … appropriately. Always drive attentively and be prepared to take immediate action" [2]. The Uber Volvo, in accident two, requires "an attentive operator to intervene if the system fails to perform appropriately" [3] and "is not designed to alert the driver that braking is needed."

As far as can be determined, the drivers in each of the cases were behind the steering wheel, but did not react in time (or at all) to the impending danger. All three drivers were not in operational control as they were not physically performing actions that led or could lead to immediate influencing of their vehicles' movement. There is a high probability that none of the pedals were in use prior to the accidents, the gearshift would not be in use, and even if the driver had their hands on the steering wheel, which we know wasn't the case in at least two of the three cases, they would not be exerting any significant force on it. In all cases, the ADCS was in operational control of the vehicles' movement, both in a longitudinal as well as lateral sense, which is by definition as the vehicles may be classified as SAE level 2 vehicles [9] with automated longitudinal and lateral control.

B. Analysis of Meaningful Human Control

As the drivers were not in operational control, we cannot automatically state that their vehicles were under MHC by only considering the driver. We need to review the design and applicable ODD of the vehicle to determine if, under the circumstances that the vehicles found themselves in, one could determine that MHC was present. The system that is considered when analyzing MHC, is defined as the complete vehicle-driver system. We therefore now consider the conditions for MHC:

a) Did the Driver-Vehicle System Act According to Some Human's Reasons to Act? (Tracking Condition)

In all three cases, the vehicles did not perform (suitable) braking maneuvers before hitting another object (i.e. pedestrian with bike, and police car). From a moral, social and legal perspective, a vehicle should arguably perform an evasive maneuver or an emergency braking maneuver if such an object is predicted to be hit. For the Tesla-trailer collision, an impaired sensor appeared to cause problems [1], while in the case of the Uber Volvo that hit the pedestrian, a false-positive was reported [3] (i.e. the pedestrian was detected, but was categorized as an anomaly and ignored), and for the Tesla and the police car, it remains unclear what the main cause of inaction by the vehicle was, although there are indications that it may have followed incorrect road markings [6], [7]. Whatever the purpose, the system did not adhere to these human reasons. Therefore, in all these cases, the system cannot be deemed to have been under MHC.

b) Is There Some Qualified Human That Can Appreciate the Capabilities of the System and Recognize Themselves as Target of Potential Moral Consequences of the System's Behavior? (Tracing Condition)

Starting with the drivers, it is unclear to what extent the drivers were sufficiently trained and knowledgeable of the system. The first part of the tracing condition should be read in a broad sense. For an agent to be able to appreciate the capabilities of a system, it is required that they have a thorough conceptual knowledge, a *know-that*, and to have the right capacities to fully and correctly utilize the system, in the sense of a *know-how*. It is reasonable to presume that the drivers trusted the system sufficiently for them to perform their monitoring tasks to the level of performance that they did. But then, the driver's capacity to perform the monitoring task was not sufficient as in each of the cases, no suitable evasive maneuvers were performed by the driver. We come back to the ability of human drivers to properly monitor an automated vehicle, as described in the introduction section. The strong evidence provided by the scientific community, and given in the introduction section, is that drivers cannot be expected to perform an engaged and active monitoring role with high situational awareness, therefore leading to high reaction times, even if this is demanded by the system or the systems designers and even by the law. Therefore, it is highly doubtful whether the drivers can properly appreciate the system's capabilities to react and appreciate their own ability to react as demanded by the ADCS (first part of the tracing condition). The second part of the tracing condition should be easier for the driver to appreciate; the moral consequence of failure to perform their role may mean the occurrence of a dangerous situation, for which they may be held responsible. This is something that the driver should be expected to appreciate, regardless of their ability to perform their role. However, as the first part of the tracing condition is not met, then the system can also not be deemed to be under MHC by the driver, based on the tracing condition. Let us then consider other potential loci of controls. Presuming no other humans were in the vehicle or exerted direct operational control through communication, then those involved in the design and production of the ADCS are arguably the only ones left that could potentially satisfy the tracing condition for MHC. While not knowing the exact algorithmic, technical, or any other design aspects of the vehicles, the vehicle manufacturers have clearly stated that their vehicles are not able to drive under all circumstances and driver monitoring is required for the other circumstances [2]. Therefore, it can be deduced that such agents do appreciate the capabilities and limitations of their automated driving system as a whole (including the human driver). However, they seem to not fully appreciate their own role as potential targets of moral blame and responsibility for possible accidents. So far, many companies, although not all, have used their disclaimers to shield themselves against legal and moral repercussions. This would lead us to conclude that designers have a limited recognition of themselves as morally responsible for behavior of the system, as prescribed by the second part of the tracing condition. This approach by manufacturers is mainly applicable to the American style legislative systems and may differ for other systems. In many countries in Europe, for example, the Vienna Convention [36] forms a main basis for the legal context, while national vehicle approval authorities play an important legal role. These approval authorities may even also be considered as targets for the tracing condition, as they play a role in approving the ADCS within the vehicle design.

Table III. Control gap in an emergency situation for a low level automated vehicle.

		Driver	ADCS	Control Exerted by…
Operational control		NO → YES (?)	YES	ADCS → Driver (?)
		Driver	ADCS Designers	Under MHC From…
MHC	Tracking	NO	NO	GAP!!!
	Tracing	NO	NO	

C. Gaps in Control

Based on the above analysis, we conclude that none of the systems exemplified by our cases could achieve meaningful human control (see Table III). In Table III, operational control is transferred to the driver (shown by the arrows), but some doubt is present to the extent that the driver actually has suitable control (hence the question marks in the table). The systems might in some sense have satisfied the tracking condition for the driver, but were judged to not satisfy tracing and possibly insufficiently able to perform the requested tasks. The tracking and the tracing conditions are very limitedly, if at all, satisfied by the designers and manufacturers of the ADCS's. As the underlying vehicle-driver system is common for partial and mixed automated vehicles (i.e. SAE level 1-3), this conclusion is also more generally valid for this level of AV. The gap in control occurs due to the vehicle being incapable of performing its tasks with its limited ODD and requiring the driver to intervene, while at the same time the driver does not have sufficient capacity to intervene. From an operational control perspective, there does not appear to be a gap in control at the moment of control transition: the vehicle can remain in control or can (partially) transfer control back to the driver regardless if the driver is ready or not. This shows that reasoning in terms of operational control, i.e. attribution of driving tasks, is far from sufficient to solve issues of "real" control and responsibility. That's why we propose to reason in terms of MHC, a notion that does not only look at the distribution of driving *tasks*, but first and foremost at the *capacities* of human and non-human agents involved in the driving operation. From the perspective of MHC, none of the potential human agents involved in the driving tasks would satisfy the "tracing" condition for the system to be deemed under control, which in turn leads to a gap in control. And a gap in control for a road vehicle has every risk of resulting in a (potentially critical) accident.

IV. Discussion and Recommendations

Much of the identified 'gaps in control' relate either directly or indirectly to the condition that the driver must 'monitor' and 'react' if required and the assumption that the driver is capable to do so. Responsibility is assigned to the driver through these conditions. However, responsibility is attributed normatively and often depends on the moral, social and legal context. Responsibility is often defined to follow control, but control cannot be attributed in the same way to responsibility. Stating that someone should be in control does not mean that they are or can be. MHC offers a concept that allows control to be attributed, checked and designed into a system in a more sensitive and encompassing way, especially when intelligent devices are involved.

When considering the design of AV's, it has already been suggested that drivers should never be disengaged from operational tasks in the first place, as mere monitoring will never suffice [11], [37]. Others have stated that levels of automation that require immediate response from drivers should never be allowed on roads and that these levels of automation should be skipped. The role and capability of drivers should be considered in AV design. Consideration of driver-vehicle control should be considered holistically and we argue from a MHC perspective that encompasses so much more than the attribution of unrealistic responsibility. And it should be noted that human drivers may not have prevented the accidents occurring either, even if they were in control. The occurrence of an accident does not have to indicate a loss of control, but can often be traced back to control gaps if they exist. For vehicle approval authorities, there is also a challenge to be able to judge the appropriateness of ADCS design. The concept of MHC also allows them to dig deeper into the control of a vehicle and set ethically acceptable and safety-conscious regulations.

The concept of MHC is still in its infancy and is still being translated into wider relevant areas of application. Specific guidelines for its use and foundations remain under development, although are expected to become more readily available in coming years. Nevertheless, MHC is already applicable in many situations regarding control, as has been shown here and in other places [34], [35], [38], [39].

V. Conclusions

In this contribution, we have demonstrated that a potential gap in vehicle-driver control exists in current partially automated vehicles that are applied in practice. This was shown by analyzing three recent and serious accidents involving partially automated AV's that serves as a proxy for many current on-road AV's, in which the driver remains in the loop. This gap in control has become evident through accidents involving automated vehicles and exists in part due to an inability of drivers to perform tasks and given responsibility. Using the concept of Meaningful Human Control (MHC), we demonstrated that although operational control might exist, control through MHC does not always exist in emergency circumstances as various circumstances fall outside of the operational design domain of many partially automated vehicles, while control (c.q. MHC) cannot always be undertaken by a driver, even if the expectation of the vehicle manufacturer and the law demand it. A recommendation is made to consider automated vehicle control from the perspective of MHC to aid a closed control system that is reasonable and humanly acceptable and achievable. This responsibility to consider vehicle control in such a way may lie with the vehicle developers and manufactures, and also with policymakers, including vehicle approval authorities.

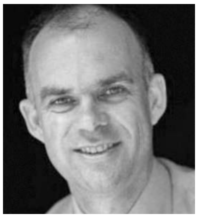
Acknowledgments

This work is part of the research project Meaningful Human Control over Automated Driving Systems with project number MVI.16.044, which is (partly) financed by the Netherlands Organization for Scientific Research (NWO).

About the Authors

Simeon C. Calvert received the M.Sc. and Ph.D. degrees in Civil Engineering, specialized in Transport & Planning form the Delft University of Technology, The Netherlands, in 2010 and 2016, respectively. He is now employed as coordinator and researcher at data and simulation lab DiTTlab at Delft University of Technology. Between 2010 and 2016 he worked as a Research Scientist at TNO, Netherlands Organization for Applied Scientific Research. There, his research has focused on ITS, impacts of vehicle automation, traffic management, traffic flow theory and network analysis. Much of his recent research has involved various roles in leading national and European research projects involving the application and impacts of vehicle automation and cooperation.

Giulio Mecacci received his M.A. in Philosophy of Mind from the University of Siena, Italy. He obtained a Ph.D. from Radboud University Nijmegen, at the Donders Institute for Brain, Cognition and Behavior, in the field of ethics of neurotechnology. He is now post-doctoral researcher at the Delft University of Technology, working together with psychologists and engineers on the multidisciplinary project "Meaningful Human Control over Automated Driving Systems". He is also assistant professor with tenure in the department of Artificial Intelligence, at Radboud University Nijmegen, dealing with ethical and societal implications of AI and intelligent technologies.



Bart van Arem received the M.Sc. and Ph.D. degrees in applied mathematics from the University of Twente, Enschede, The Netherlands, in 1986 and 1990, respectively. From 1992 and 2009, he was a Researcher and a Program Manager with TNO, working on intelligent transport systems. Since 2009, he has been the Chair Professor of Transport Modeling with the Department of Transport and Planning, Delft University of Technology, Delft, The Netherlands, focusing on the impact of intelligent transport systems on mobility. His research interests include transport modelling and intelligent vehicle systems. In 2017, he received the IEEE ITS Society Institutional Lead Award for the TU Delft Automated Driving Research Program.

Filippo Santoni de Sio received the Ph.D. in Ethics and Philosophy of Law from the University of Turin, Italy in 2008. From late 2008 to late 2011, he worked as a Post-doc at the University of Turin on a project entitled 'Irresistible Desires as an Excuse' aimed to answer the problem of moral and legal responsibility of offenders affected with compulsive mental disorders or addiction. From 2012 to 2014, he was engaged as a post-doc in an international and interdisciplinary project entitled 'Enhancing responsibility: the effects of cognitive enhancement on moral and legal responsibility'. He is currently Assistant Professor in Philosophy at TU Delft with a main research focus on problems of personal responsibility and meaningful human control. He is also a member of the Dutch 4TU task force for robotics.

Daniël D. Heikoop obtained his M.Sc. in Applied Cognitive Psychology at the University Utrecht, after which he started a Ph.D. within the Marie Curie-Skłodowska Actions funded project called HFAuto. Between 2014 and 2017 he performed his Ph.D. on Driver Psychology during Automated Platooning as an external student from Delft University of Technology at the University of Southampton (UK). He now works at the Delft University of Technology, on the project called "Meaningful Human Control over Automated Driving Systems", in which he actively collaborates with psychologists, traffic engineers, and philosophers.

Marjan Hagenzieker received M.Sc. and Ph.D. degrees in psychology and social sciences from the University of Leiden, The Netherlands, in 1987 and 1999, respectively. From 1987 to 2018 she was researcher and scientific advisor at SWOV Institute for Road Safety Research in the Netherlands. Since 2014, she is full Professor Traffic Safety with the Department of Transport and Planning, Delft University of Technology, Delft, The Netherlands. Her current research focuses on how to ensure road safety in modern urban environments with many kinds of road users and divergent interests. Specific research topics include the behavior and safety of vulnerable road users, and road user interactions with road infrastructure, in-vehicle technology, and automated vehicles.

References

[1] "Office of Defective Investigation – preliminary report summary 2015 Tesla Model S Crash of May 7, 2016," DoT – National Highway Traffic Safety Administration, Washington, D.C., 2016.
[2] *Model S Owner's Manual*, vol. 31. Tesla, 2016, p. 2016.
[3] "Preliminary report highway HWY18MH010 – Uber Volvo XC90 incident March 18, 2018," National Transportation Safety Board, Washington, D.C., 2018.
[4] F. M. Favarò, N. Nader, S. O. Eurich, M. Tripp, and N. Varadaraju, "Examining accident reports involving autonomous vehicles in California," *PLoS ONE*, vol. 12, p. e0184952, 2017. doi: 10.1371/journal.pone.0184952.
[5] N. Kalra, "Challenges and approaches to realizing autonomous vehicle safety," 2017.
[6] R. Stumpf, "Tesla on autopilot crashes into parked California police cruiser," *The Drive*, June 18, 2018. [Online]. Available: www.thedrive.com/news/21172/tesla-on-autopilot-crashes-into-parked-california-police-cruiser
[7] AP, "Tesla in autopilot mode crashes into California police car," June 18, 2018. [Online]. Available: www.apnews.com/47e78d649678424f97569062d140a2c2/Tesla-in-Autopilot-mode-crashes-into-California-police-car
[8] I. Y. Noy, D. Shinar, and W. J. Horrey, "Automated driving: Safety blind spots," *Safety Sci.*, vol. 102, pp. 68–78, 2018. doi: 10.1016/j.ssci.2017.07.018.
[9] *Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems*, SAE J3016, 2018.
[10] J. Axelsson, "Safety in vehicle platooning: A systematic literature review," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, pp. 1033–1045, 2017. doi: 10.1109/TITS.2016.2598873.
[11] S. M. Casner, E. L. Hutchins, and D. Norman, "The challenges of partially automated driving," *Commun. ACM*, vol. 59, pp. 70–77, 2016. doi: 10.1145/2830565.
[12] N. Merat, A. H. Jamson, F. C. Lai, M. Daly, and O. M. Carsten, "Transition to manual: Driver behaviour when resuming control from a highly automated vehicle," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 27, pp. 274–282, 2014. doi: 10.1016/j.trf.2014.09.005.
[13] J. C. De Winter, R. Happee, M. H. Martens, and N. A. Stanton, "Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 27, pp. 196–217, 2014. doi: 10.1016/j.trf.2014.06.016.
[14] A. H. Jamson, N. Merat, O. M. Carsten, and F. C. Lai, "Behavioural changes in drivers experiencing highly-automated vehicle control in varying traffic conditions," *Transp. Res. C, Emerg. Technol.*, vol. 30, pp. 116–125, 2013. doi: 10.1016/j.trc.2013.02.008.
[15] N. Merat, A. H. Jamson, F. C. Lai, and O. Carsten, "Highly automated driving, secondary task performance, and driver state," *Hum. Factors*, vol. 54, pp. 762–771, 2012. doi: 10.1177/0018720812442087.
[16] M. S. Young and N. A. Stanton, "Malleable attentional resources theory: A new explanation for the effects of mental underload on performance," *Hum. Factors*, vol. 44, pp. 365–375, 2002. doi: 10.1518/0018720024497709.
[17] M. Karashima and M. Saito, "A study on the error occurrence and human information processing time influenced by the fluctuation of working memory resource capacity," *Int. J. Cogn. Ergon.*, vol. 5, pp. 91–109, 2001. doi: 10.1207/S15327566IJCE0502_1.
[18] R. Fuller, "Towards a general theory of driver behaviour," *Accid. Anal. Prev.*, vol. 37, pp. 461–472, 2005. doi: 10.1016/j.aap.2004.11.003.
[19] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," *Hum. Factors*, vol. 37, pp. 32–64, 1995. doi: 10.1518/001872095779049543.
[20] M. R. Endsley, *Designing for Situation Awareness: An Approach to User-Centered Design*. Boca Raton, FL: CRC, 2016.

[21] D. D. Heikoop, J. C. de Winter, B. van Arem, and N. A. Stanton, "Psychological constructs in driving automation: A consensus model and critical comment on construct proliferation," *Theor. Issues Ergonom. Sci.*, vol. 17, pp. 284–303, 2016. doi: 10.1080/1463922X.2015.1101507.
[22] C. C. Jacoby, and S. K.Schuster, "Issues of automated vehicles operating in mixed traffic," in *Proc. IEEE Conf. Intelligent Transportation System (ITSC'97)*, 1997, pp. 607–612. doi: 10.1109/ITSC.1997.660543.
[23] T. B. Sheridan, *Humans and Automation: System Design and Research Issues*. Human Factors and Ergonomics Society, 2002.
[24] N. Strand, J. Nilsson, I. M. Karlsson, and L. Nilsson, "Semi-automated versus highly automated driving in critical situations caused by automation failures," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 27, pp. 218–228, 2014. doi: 10.1016/j.trf.2014.04.005.
[25] J. F. Mackworth, "Performance decrement in vigilance, threshold, and high-speed perceptual motor tasks," *Can. J. Psychol.*, vol. 18, p. 209, 1964. doi: 10.1037/h0083302.
[26] A. Eriksson and N. A. Stanton, "Driving performance after self-regulated control transitions in highly automated vehicles," *Hum. Factors*, vol. 59, pp. 1233–1248, 2017. doi: 10.1177/0018720817728774.
[27] W. P. Vlakveld, "Transition of control in highly automated vehicles: A literature review," 2016.
[28] R. Happee, C. Gold, J. Radlmayr, S. Hergeth, and K. Bengler, "Take-over performance in evasive manoeuvres," *Accid. Anal. Prev.*, vol. 106, pp. 211–222, 2017. doi: 10.1016/j.aap.2017.04.017.
[29] C. Gold, D. Damböck, L. Lorenz, and K. Bengler, "Take over!" How long does it take to get the driver back into the loop?" in *Proc. Human Factors and Ergonomics Society Annu. Meeting*, 2013, pp. 1938–1942. doi: 10.1177/1541931213571433.
[30] J. Radlmayr, C. Gold, L. Lorenz, M. Farid, and K. Bengler, "How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving," in *Proc. Human Factors and Ergonomics Society Annu. Meeting*, 2014, pp. 2063–2067. doi: 10.1177/1541931214581434.
[31] Z. Lu, R. Happee, C. D. Cabrall, M. Kyriakidis, and J. C. de Winter, "Human factors of transitions in automated driving: A general framework and literature survey," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 43, pp. 183–198, 2016. doi: 10.1016/j.trf.2016.10.007.
[32] J. A. Michon, "A critical view of driver behavior models: What do we know, what should we do?" in *Human Behavior and Traffic Safety*. New York: Springer-Verlag, 1985, pp. 485–524.
[33] S. C. Calvert, D. D. Heikoop, and B. Van Arem, "Core components framework of automated driving systems with meaningful human control," submitted for publication.
[34] F. Santoni de Sio and J. Van den Hoven, "Meaningful human control over autonomous systems: A philosophical account," Frontiers Robotics AI, 2018. doi: 10.3389/frobt.2018.00015.
[35] F. Santoni de Sio, "Ethics and self-driving cars: A white paper on responsible innovation in automated driving systems," Dutch Ministry Infrastructure Environment Rijkswaterstaat, 2016.
[36] Economic Commission for Europe, "Convention on road traffic," *Inland Transport Committee*, Vienna, Nov. 8, 1968.
[37] N. Merat and J. D. Lee, "Preface to the special section on human factors and automation in vehicles: Designing highly automated vehicles with the driver in mind," *Hum. Factors*, vol. 54, pp. 681–686, 2012. doi: 10.1177/0018720812461374.
[38] M. C. Horowitz and P. Scharre, "Meaningful human control in weapon systems," 2015.
[39] T. Marauhn, "Meaningful human control – and the Politics of International Law," in *Dehumanization of Warfare*. New York: Springer-Verlag, 2018, pp. 207–218.

ITS

Keywords—Vehicle control; Meaningful human control; Vehicle automation; Autonomous driving