

Keeping the Driver in the Loop: The ‘Other’ Ethics of Automation

Victoria Banks¹[0000-0001-9623-5805], Emily Shaw², David R. Large²[0000-0003-3046-4984]

¹ Human Factors Engineering, Transportation Research Group, Southampton, UK

² Human Factors Research Group, University of Nottingham, Nottingham, UK
david.r.large@nottingham.ac.uk

Abstract. Automated vehicles are expected to revolutionise everyday travel with anticipated benefits of improved road safety, comfort and mobility. However, they also raise complex ethical challenges. Ethical debates have primarily centred around moral judgements that must be made by autonomous vehicles in safety-critical situations, with proposed solutions typically based on deontological principles or consequentialism. However, ethics should also be acknowledged in the design, development and deployment of partially-automated systems that invariably rely upon the human driver to monitor and intervene when required, even though they may be ill-prepared to do so. In this literature review, we explore the lesser-discussed ethics associated with the role of, and expectations placed upon, the human driver in partially-automated vehicles, discussing factors such as the marketing and deployment of these vehicles, and the impact upon the human driver’s development of trust and complacency in automated functionality, concluding that the human driver must be kept ‘in the loop’ at all times.

Keywords: Ethics, Automated Driving, Literature Review.

1 Introduction

Automated driving is expected to enhance road safety by reducing the number of accidents attributable to driver error, improve comfort by reducing driver workload and enabling drivers to engage in secondary tasks and activities, and deliver ‘mobility for all’ [1]. These are undeniably bold and admirable claims, which on face value suggest that the current drive towards the ‘driverless car’ favours the ‘greater good’ and should be applauded and encouraged. However, the visionary hype surrounding the concepts of ‘self-driving’, ‘driverless’ and ‘autonomous cars’ is masking a disconnect between public expectations and current technological capabilities [2,3]. Journalists in the popular media are failing to clarify important differences in terminology that define the varying levels of automation within the automotive industry. Vehicle manufacturers are also using choice words to describe product capabilities that carry a multitude of interpretations [4]. For example, it has been suggested that Tesla’s choice of the name ‘Autopilot’ for its partially automated driving feature suggests that it has fully autonomous capability and therefore, promises more than it can deliver [5,6]. Indeed, Tesla faced

intense media scrutiny following the first fatal collision involving one of its vehicles being operated in Autopilot mode in May 2016. The driver of a Tesla Model S was fatally injured when their vehicle collided with a tractor trailer that was crossing an intersection on a highway west of Williston, Florida. Data taken directly from the Tesla Model S vehicle in question confirmed that, at the time of the incident, the vehicle was being operated in Autopilot mode. The Autonomous Emergency Brake (AEB) system did not provide any warning nor initiated automated braking and there was no attempt made by the driver to take evasive action. There was speculation that the radar and camera technology used on Tesla's AEB system failed to detect the trailer against a brightly lit sky, or the trailer was misclassified as an overhead sign by the software. However, the National Highway Traffic Safety Administration [7] concluded that "human error" was the primary cause of the fatal collision and speculated that the driver must have been distracted from the driving task for an 'extended period'. Further investigation by the National Transportation Safety Board [8] reported that the drivers' hands were only detected for 25 seconds over a thirty-seven minute journey. It is therefore reasonable to assume that the driver was on the whole operating the vehicle completely "hands and feet free", despite Tesla's own Human-Machine Interface (HMI) alert that states "Always keep your hands on the wheel, be prepared to take over at any time" when Autopilot is initially activated. According to Norman [9,10], automation is at its most dangerous when it behaves in a consistent and reliable manner for most of the time. This is because of an increased risk of out-of-the-loop performance problems including complacency [11], over-trust [12], and loss of situation awareness [13].

Instead, the full benefits of automation can only be realised if the driver is completely removed from the driving task. In these situations, the vehicle would then need to make moral and ethical judgements in otherwise conflicting safety-critical situations. The development of control algorithms and artificial intelligence to enable such decisions is therefore the subject of many ongoing works. Studies propose the use of either: deontological principles, such as the laws of robotics [14]; or consequentialism, which uses the classic 'trolley dilemma', a thought-scenario based on unavoidable harm, to determine moral right and wrong in order to inform computational algorithms [15,16,17]. However, we currently reside in an era whereby only some driving tasks are automated. This means that the human driver still plays a vital role within the wider driving task.

1.1 The Role of Terminology

One of the greatest challenges surrounding automobile automation is the inconsistent use of terminologies, in particular, two key concepts: autonomous and automation. Understanding the differences between these terms is extremely important because the terminologies used to describe a feature can lead to inferences over what the role of the human driver is within the wider system. According to the Oxford English Dictionary [18], to be 'autonomous' is to mean "having the freedom to act independently". An autonomous vehicle therefore can act independently and control its own behaviour without human intervention. One example of an 'autonomous' vehicle system is the Electronic Stability Control (ESC) system that can improve vehicle stability by detecting and then reducing a loss of traction that can lead to skidding.

In contrast, ‘automation’ refers to the “use or introduction of automation equipment” [18]. Thus, an automated vehicle is not capable of acting independently from the driver, and any ‘automated control’ must be carefully engineered to ensure seamless transition from and back to manual control. The driver therefore has overarching control of vehicle operation meaning that automated systems can be disengaged whenever the driver feels it is necessary. Automated systems can therefore be considered as ‘driver-initiated’ because it is the human that has ultimate authority of the primary task.

There are also various levels of automated functionality that further complicate our understanding of automation. The Society of Automotive Engineers (SAE) [19] defined five levels of automation specific to the driving domain, outlining the functional capabilities of systems operating within each of these levels. SAE propose that automation ranges from Level 0 (Fully Manually) to Level 5 (Fully Autonomous). However, this can be criticised because automation is not considered to be a dichotomous concept. Instead, automation operates on a continuum [20] making it difficult to classify systems whereby different component parts operate at different levels of automation [2].

There are, however, a number of taxonomies that have sought to better define the role of the human within automated systems. The oldest and most widely cited taxonomy was developed by Sheridan and Verplanck [21], who offer a ten-level taxonomy specifying which functions are the responsibility of the human operator and which are the responsibility of the computer system. Endsley and Kaber [22] sought to better define the intermediate levels by identifying “who” was doing “what” in terms of system monitoring, strategy generation, decision making and response execution processes. Defining the allocation of system function in this way can better help us understand the explicit roles and responsibilities of the driver at each level of automation.

1.2 The Role of the Driver

Kaber and Endsley [23] captured the idea of the changing driver role eloquently by describing a shift from ‘active operation’ to ‘passive monitoring’ as the level of automation increases. However, the expectation on the human driver to passively monitor operations and rapidly resume control in emergency situations, or when systems go beyond their parameters, is posited as the main human factors challenge facing the deployment of automated vehicles [2]. It is argued that the designers of automated systems have a moral and ethical responsibility to consider how the operational characteristics and capabilities of the systems impact upon the human driver. Poulin et al. [24] go as far to suggest that drivers will end up with none of the control but all of the accountability for a system that has not been designed in recognition of any potential performance issues. This is termed “the responsibility gap”, and is often an inevitable outcome relating to the complexity of a system [25]. A responsibility gap would be the consequence of technologies being deployed with little knowledge as to how they will behave in context. In the case of autonomy, this would be a result of autonomous vehicles being programmed to learn as they operate. This would also mean that not even the human initially responsible for programming these vehicles, would be able to understand or predict the processes underpinning artificial agent decision making in future scenarios [26].

Too much automation can take the human out of the loop, deskill them and lower morale [27, 28]. In addition, automated assistance can lead to decrements in situation awareness and in some instances cause erratic changes to driver mental workload [29]. This is because any sudden transfer of control may result in mode confusion [30] and/or startle effects [31]. Finding ways to encourage and motivate drivers to stay in-the-loop is an enduring challenge, one that could rely heavily upon the role of the driver monitoring. Hancock [32], however, suggests that there is a need to consider motivation theory in relation to human-computer interaction associated with the automotive domain. The general consensus is that mental workload optimisation is crucial in maintaining effective task performance (e.g. [33]). Thus, being able to recognise different driver states at higher levels of automation will enable researchers to not only explore the most efficient strategies of keeping the driver ‘in-the-loop’ but always the most efficient means to transfer control.

The application of automation by designers without due consideration of the consequences to human performance is something Parasuraman and Riley [34] referred to as ‘abuse’ of automation and arguably violates the ethical principle of respect for persons [35, p1845]. Without a human-centred focus, the benefits of automation for individual drivers are overlooked in pursuit of higher levels of functionality [36]. We also miss opportunities to learn from driver experiences in using such systems that could help inform future generations of technology that can improve the safety of both its users and wider public [6].

At the core of the Human Factors and Ergonomics discipline is the goal to enable the development of innovative and beneficial technologies that can optimise the performance and safety of the human that interact with them [36]. Systems engineering uses the principles of complementarity to help inform the design of complex sociotechnical systems such as those seen in driving automation. This approach is based upon the idea that the allocation of tasks should serve to maintain control whilst retaining human skill [37]. Such an approach aims to encourage shared situation awareness between driver and automated systems [38]. ‘Team cognition’ appears to be the binding mechanism that goes on to produce coordinated behaviour [39]. According to Cuevas et al. [39], a human-automation team can be defined as the coupling of both human and automated systems working both collaboratively and in coordination with one another to successfully complete a task. Hoc [40] originally applied the concept of cooperation as a means to identify, analyse, implement and support cooperative activities between humans and machines. This essentially describes the essence of partially automated driving solutions, but does not necessarily reflect the reality of such systems.

1.3 The Role of Marketing

The drive for vehicle automation is likely to continue at pace given the potential economic benefits available from competitive advantage within a mass-market commodity [32, 41]. However, this enthusiasm and push for technological innovation may encourage more focus on the functional capabilities and characteristics of a system rather than the impacts that it may have on driver behaviour. It is after all the former that dominates discussions within the public arena [32]. The marketing and deployment of automated

technologies may therefore negatively impact upon driver expectations [6, 42], risk perception [43] and trust [11]. All of these factors can contribute to automation ‘misuse’ (i.e. over reliance) and/or ‘disuse’ (i.e. underutilisation) [34].

The marketing of automated systems is also likely to influence how drivers interpret and predict what may happen within the driving environment (i.e. their mental model of the system) [3,27], particularly if these are reinforced by positive interactions with technology (i.e. systems appears to be reliable, accurate and dependable). Of course, any deviation in expected behaviour has the potential to challenge existing models and affect user trust in system functionality [44].

There seems to be a lack of transparency regarding the true technological capabilities of autonomous systems. For example, Endsley [45] recalls that a service representative from Tesla initially told her the automated systems were 100% reliable. However, when it comes down to a matter of liability, Tesla were clear to state that the Autopilot in use around the time of the fatal accident in May 2016 was in the beta testing phase [46]. Mobileye, who make the image recognition software for assisted driving, condemned Tesla’s claims that their Autopilot was safer than a human driver on highways by citing the dangers involved with causing mistrust through consumer confusion of technological capabilities [47]. In fact, Smith [48] calculated that in order to say with 99% confidence that automated vehicles crash less frequently than vehicles with human drivers, it would need to drive 725,000 mile on representative roadways without incident. If only fatal crashes were considered, this figure rises dramatically to 300 million miles. To date, no automated vehicle has yet to travel such distances unassisted [49] and therefore it is not possible to validate the claim that automated vehicles are safer than vehicles driven by human drivers. In reality, even with perfect sensing, it is unrealistic to suggest that a truly crash free environment can exist [49]. This is because the driving environment is characterised by unpredictable obstacles and dynamically changing situations that are difficult to design for. We also cannot overlook the fact that inevitable failures will occur. Software and/or hardware failures remain an inevitable threat to the safety of automated systems. In an effort to minimise such failures, vehicles will need multiple redundancies, extensive testing and likely need mandatory maintenance [49].

1.4 The Role of Technology

Given that software and/or hardware failures appear to remain an inevitable threat to drivers [49], research into failure-induced transfer of control has been extensively studied [50, 51]. Early research concentrated on determining appropriate thresholds relating to how long a driver would need to know in advance about an upcoming control transition. Damböck et al. [52] utilised three different lead times – 4, 6 and 8 seconds – and found that performance did not differ significantly from manual driving when an 8 second lead time was used. Similarly, Gold et al. [53] suggested that drivers needed a lead time of 7 seconds to ensure a safe resumption of control. For partially automated driving, the literature suggests that drivers can take anywhere between 1.2 seconds [54] and 15 seconds to respond to emergency situations [55]. Further, for self-paced control transitions (i.e. non-critical), response times vary between 1.9 and 25.7 seconds depending upon task engagement and criticality [56]. This has significant implications for

the design of warning systems as they must provide drivers with adequate warning about takeover situations. Whilst it is common practice to design for the 90th percentile (taking into the consideration the range between the 5th percentile female and 95th percentile male populations [57]), it is important that manufacturers are given more information about the ‘extreme’ ends of the population (i.e. range in performance). Using only mean values as a basis to design systems is troublesome because we do not understand the range or spread of distributions. Thus, Eriksson and Stanton [56] argue that a more inclusive design approach is needed. Median response times are simply not sufficient enough when it comes to designing control transitions for automated driving.

Even so, Louw [58] argues that, in addition to providing adequate warning to drivers, a more pressing need is to develop an objective measure relating to handover quality and safety. Thomas et al. [59] suggest that it is in the seconds immediately after takeover when most errors occur. For the case of partial automation, drivers can take between 35 and 40 seconds to stabilise lateral control of the vehicle following a transfer of control from automated to manual driving [55]. Transfer of control and the provision of adequate warning remains an ongoing challenge.

2 Conclusions

The utilitarian justification for automated vehicles is that they can improve safety, comfort and deliver greater mobility for both the user and wider population [1, 2, 47]. However, there is a need to balance these justifications with the current state of the art. In reality, fully autonomous vehicles are a long way off [10]. Partially automated driving solutions are already available today, but can give the impression of higher level functionality. Systems that can automate both longitudinal and lateral control, as well as automate aspects of traditional driver decision-making enable drivers to become “hands and feet free”. However, they still require the driver to supervise, maintain and potentially take over from the system [2, 10, 45]. Unfortunately, humans are notoriously inefficient at completely sustained vigilance tasks (e.g. [50, 60]). Combine that with the fact that drivers can quickly become complacent in situations where automated systems behave in a consistent manner for extended periods, it appears inevitable that drivers will over-trust systems offering limited automated assistance (e.g. [11, 61, 62]). This behaviour will heavily impact upon the capabilities of the driver to respond appropriately when required.

Whilst the concepts of “automation” and “autonomy” are clearly distinct, it is worth noting that some of the more sophisticated *automated* technologies have the potential to become *autonomous* in the future [32]. This is of course a natural progression for technological development. For example, at a basic level ‘cruise assist’ technologies have evolved from systems that simply maintain a pre-set speed (Cruise Control), to systems that can adapt their speed profile depending upon the speed of the vehicle ahead, although notably will not exceed a pre-set maximum (Adaptive Cruise Control). It therefore seems likely that continued extension of pre-existing automated systems could pave the way for future autonomous functionality. Presently, however, it appears

that we do not fully understand or appreciate the complexities of human-machine interaction in vehicles offering enhanced automated driving solutions [63]. This is further exacerbated by the current approach to accident investigation following an incident involving an ‘automated’ vehicle. Rather than adopting a holistic ‘system’ view, individual subsystems tend to be looked at in isolation; if they are deemed to be operationally sound, blame will often be attributed to ‘human error’ [64]. Such an approach is problematic because it fails to consider how multiple subsystems operating together can synergistically change the way in which the vehicle performs and the impact it has on the driver. It also fails to consider contextual factors that could significantly influence the operator’s behaviour within the system (e.g. how automated systems are named and marketed). If the circle of investigation widens to include such factors, it is likely that the conclusions drawn from the investigation process will provide a much more accurate and informative depiction of accident causation. After all, Dekker [65] argues that human error should be the starting point of an accident rather than the end, and that any case involving human error “demands an explanation” [65, p.68].

References

1. Stanton, N. A., Marsden, P. 1996. From fly-by-wire to drive-by-wire: Safety implications of automation in vehicles. *Safety Science* 24(1): 35–49.
2. Kyriakidis, M., De Winter, J. C. F., Stanton, N., Bellet, T., Van Arem, B., Brookhuis, K., Martens, M. H., Bengler, K., Andersson, J., Merat, N., Reed, N., Flament, M., Hagenzieker, M., Happee R. 2016. A human factors perspective on automated driving. *Theoretical Issues in Ergonomics Science*, pp.1-27
3. Neilsen, J. 2010. Mental models. Available at <https://www.nngroup.com/articles/mental-models/>. Accessed 15.08.2017.
4. Shladover, S. E. 2016. The truth about self-driving cars. *Scientific American*, 314(6), 52-57.
5. Abraham, H., Seppelt, B., Mehler, B., Reimer, B. 2017. What’s in a name: Vehicle technology branding consumer expectation for automation. *Proceedings of AutomotiveUI ’17*, Oldenburg, Germany.
6. Stilgoe, J. 2017. Tesla crash report blames human error – This is a missed opportunity. *The Guardian*. Retrieved from <https://www.theguardian.com/science/political-science/2017/jan/21/tesla-crash-report-blames-human-error-this-is-a-missed-opportunity>. Accessed 14 Aug 2017.
7. National Highway Traffic Safety Administration. 2017. ODI Resume. Available at <https://static.nhtsa.gov/odi/inv/2016/INCLA-PE16007-7876.PDF>. Accessed 15.08.2017.
8. National Transportation Safety Board. 2017. Driver assistance systems specialists factual report. Available at: <https://dms.nts.gov/pubdms/>. Accessed 09.08.2017.
9. Norman, D. A. 1990. The “problem” with automation: inappropriate feedback and interaction, not “over- automation”. *Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences*, 1327(1241), 585–93.
10. Norman, D. A. 2015. The human side of automation. In *Road Vehicle Automation 2* (pp. 73-79). Springer International Publishing.
11. Lee, J. D. See, K. A. 2004. Trust in automation: Designing for appropriate reliance. *Human Factors*, 46, 50– 80.

12. Walker, G. H., Stanton, N. A., Salmon, P. M. 2016. Trust in vehicle technology. *International Journal of Vehicle Design*, 70(2), 157-182.
13. Endsley, M. R. 1995. Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1), 32- 64.
14. Asimov, I. 1942. *I Robot*. New York: Gnome Press.
15. Murphy, R. R., Woods, D. D. 2009. Beyond Asimov: The three laws of responsible robotics. *IEEE Intelligent Systems*, 24(4), 14-20.
16. Stüttgen, L.R., Gast, R., König, P., Pipa, G. 2017. Using virtual reality to assess ethical decisions in road traffic scenarios: applicability of value-of-life-based models and influences of time pressure. *Frontiers in Behavioral Neuroscience*, 11, 122.
17. Skulmowski, A., Bunge, A., Kaspar, K., Pipa, G. 2014. Forced-choice decision-making in modified trolley dilemma situations: a virtual reality and eye tracking study. *Frontiers in Behavioral Neuroscience*, 8.
18. Oxford English Dictionary. 2017. Available at <https://en.oxforddictionaries.com/definition/autonomous>. Accessed 15.08.2017.
19. Smith, B. W. 2012. Driving at perfection. The Center for Internet and Society at Stanford Law School, March 2012. <http://cyberlaw.stanford.edu/blog/2012/03/driving-perfection>. Accessed 15.08.2017.
20. Parasuraman, R, Sheridan, T. B., Wickens, C. D. 2000. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics. Part A, Systems and humans: A publication of the IEEE Systems, Man, and Cybernetics Society*, 30(3), 286-97.
21. Sheridan, T.B., Verplanck, W.L. 1978. *Human and Computer Control of Undersea Teleoperators*. MIT Man-Machine Laboratory, Cambridge MA.
22. 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.
23. 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.
24. Poulin, C., Stanton, N. A., Cebon, D., Epple, W. 2015. Responses to Autonomous Vehicles. *Ingenia* 62: 8– 11.
25. Matthias, A. 2004. The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, 6(3), 175-183.
26. Johnson, D. G., Norman, M. 2014. Recommendations for future development of artificial agents. *IEEE Technology and Society Magazine*, Winter 2014, 22-28.
27. Stanton, N. A., Young, M. S., Walker, G. H. 2007. The psychology of driving automation: a discussion with Professor Don Norman. *International Journal of Vehicle Design*, 45(3), 289-306.
28. Bainbridge, L. 1983. Ironies of automation. *Automatica*, 19(6), 775-779.
29. Young, M. S., Stanton, N. A. 2002. Malleable attentional resources theory: A new explanation for the effects of mental underload on performance. *Human Factors*, 44(3), 365-375.
30. Sarter, N. B., Woods, D. D., Billings, C. E. 1997. Automation surprises. *Handbook of Human Factors and Ergonomics*, 2, 1926-1943.

31. Sarter, N. B., Woods, D. D. 1995. How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors*, 37, 5–19.
32. Hancock, P. A. 2016. Imposing limits on autonomous systems. *Ergonomics*, 60(2), 284-291.
33. Wilson, J. R., Rajan, J. A. 1995. Human-machine interfaces for systems control. In J. R. Wilson E. N. Corlett (Eds.), *Evaluation of human work: a practical ergonomics methodology*. (pp. 357-405). London: Taylor Francis.
34. Parasuraman, R., Riley, V. 1997. Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230-253.
35. Kumfer, W. J., Levulis, S. J., Olson, M. D., Burgess, R. A. 2016. A human factors perspective on ethical concerns of vehicle automation. *Human Factors*, 60(1), 1844-1848.
36. Hancock, P. A. 2014. Automation: how much is too much? *Ergonomics*, 57(3), 449-454.
37. Grote, G., Weik, S., Wafler, T., Zolch, M., 1995. Criteria for the complementary allocation of functions in automated work systems and their use in simultaneous engineering projects. *International Journal of Industrial Ergonomics* 16, 326–382.
38. Stanton, N. A., Stewart, R., Harris, D., Houghton, R. J., Baber, C., McMaster, R., Salmon, P. M., Hoyle, G., Walker, G., Young, M. S., Linsell, M., Dymott, R., Green, D. 2006. Distributed situation awareness in dynamic systems: theoretical development and application of an ergonomics methodology. *Ergonomics*, 49(12- 13), 1288–1311.
39. Cuevas, H. M., Fiore, S. M., Caldwell, B.S., Strater, L. 2007. Augmenting team cognition in human– automation teams performing in complex operational environments. *Aviation, Space, and Environmental Medicine*, 78, B63–B70.
40. Hoc, J. M. 2000. From human-machine interaction to human-machine cooperation. *Ergonomics*, 43(7), 833- 843.
41. Parasuraman, R., & Wickens, C. D. 2008. Humans: still vital after all these years of automation. *Human Factors*, 50, 511-520.
42. Merat, N., Lee, J. D. 2012. Preface to the special section on human factors and automation in vehicles designing highly automated vehicles with the driver in mind. *Human Factors*, 54(5), 681-686.
43. Brill, J. C., Bliss, J. P., Hancock, P. A., Manzey, D., Meyer, J., Vredenburg, A. 2016. Matters of ethics, trust, and potential liability for autonomous systems. *Human Factors*, 60(1), 308-312.
44. Banks, V. A., Stanton, N. A. 2015. Discovering driver-vehicle coordination problems in future automated control systems: Evidence from verbal commentaries. *Procedia Manufacturing*, 3, 2497-2504.
45. Endsley, M. R. 2017. Autonomous Driving Systems: A Preliminary Naturalistic Study of the Tesla Model S. *Journal of Cognitive Engineering and Decision Making*.
46. Tesla Motors 2016. A Tragic Loss. Tesla (press release). Retrieved from <https://www.tesla.com/blog/tragic-loss>. Accessed 15.08.2017.
47. Brown, B., Laurier, E. 2017. The trouble with autopilots: Assisted and autonomous driving on the social road. *Proceedings of Human Factors in Computing Systems*.
48. Society of Automotive Engineers. 2016. Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems. Available at: http://standards.sae.org/j3016_201401/. Accessed 12/10/2015.

49. Goodall, N. J. 2014. Ethical decision making during automated vehicle crashes. *Transportation Research Board*, 2424, 58-65.
50. Molloy, R., Parasuraman, R. 1996. Monitoring an automated system for a single failure: Vigilance and task complexity effects. *Human Factors*, 38, 311-322.
51. Strand, N., Nilsson, J., Karlsson, I. C. M., Nilsson, L. 2014. Semi-automated versus highly automated driving in critical situations caused by automation failures. *Transportation Research Part F: Traffic Psychology and Behaviour*, 27(Part B), 218-228.
52. Damböck, D., Bengler, K., Farid, M., Tönert, L. 2012. Übernahmezeiten beim hochautomatisierten Fahren [Takeover times for highly automated driving]. *Tagung Fahrerassistenz*, 15, 16-28.
53. Gold, C., Damböck, D., Lorenz, L., Bengler, K. 2013. 'Take over!' How long does it take to get the driver back into the loop? In *Proceedings of the Human Factors and Ergonomics Society 57th Annual Meeting* (pp. 1938-1942). Santa Monica, CA: Human Factors and Ergonomics Society.
54. Zeeb, K., Buchner, A., Schrauf, M. 2015. What determines the take-over time? An integrated model approach of driver takeover after automated driving. *Accident Analysis Prevention*, 78, 212-221.
55. Merat, N., Jamson, A. H., Lai, F. C. H., Daly, M., Carsten, O. M. J. 2014. Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour*, 26(Part A), 1-9.
56. Eriksson, A., Stanton, N. A. 2017. Take-over time in highly automated vehicles: Non-critical transitions to and from manual control. *Human Factors*, 59(4), 689-705.
57. Porter, J. M., Case, K., Marshall, R., Gyi, D., Sims neé Oliver, R. 2004. 'Beyond Jack and Jill': designing for individuals using HADRIAN. *International Journal of Industrial Ergonomics*, vol 33, no 3, pp 249-264.
58. Louw, T., Merat, N., Jamson, H. 2015, June. Engaging with highly automated driving: To be or not to be in the loop? *Eighth International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, Salt Lake City, Utah.
59. Thomas, M. J., Schultz, T. J., Hannaford, N., Runciman, W. B. 2013. Failures in transition: Learning from incidents relating to clinical handover in acute care. *Journal for Healthcare Quality*, 35(3), 49-56.
60. Casner S. M., Schooler, J. W. 2015. Vigilance impossible: Diligence, distraction, and day-dreaming all lead to failures in a practical monitoring task. *Consciousness and Cognition*, 35, pp 33-41.
61. Parasuraman, R., Molloy, R., Singh, I. L. 1993. Performance consequences of automation-induced 'complacency'. *The International Journal of Aviation Psychology*, 3, 1-23.
62. Hollnagel, E., Woods, D. D. 2005. *Joint Cognitive Systems Foundations of Cognitive Systems Engineering*, CRC Press.
63. Weyer, J., Fink, D., Adelt, F. 2015. Human-machine cooperation in smart cars. An empirical investigation of the loss-of-control thesis. *Safety Science*, 72, 199-208.
64. Rasmussen, J. 1990. Human error and the problem of causality in analysis of accidents. *Philosophical Transactions of the Royal Society London*, 327, 449-462.
65. Dekker, S. W. 2006. *The Field Guide to Understanding Human Error*. Aldershot, Ashgate.