1.2.1 – Considering human/machine interactions

Practical guidance – automotive

Authors: Professor Robin E Bloomfield, Gareth Fletcher, and Heidy Khlaaf (Adelard LLP)

Roads are "inherently social" in nature [9]. Autonomous Vehicles (AVs) will share the road with many other users (pedestrians, cyclists, emergency vehicles, construction crews, human-operated cars, etc.). There is, therefore, an argument for modelling social aspects when developing technology [1][2] - going beyond technology and logistics, which tend to be the focus of safety analyses [3]. Inevitably, an AV's actions will have effects on what human-operated vehicles will do, as discussed in [4], for example.

There is a range of concerns that involve the relationship between people and AVs. At one end of the spectrum, there are basic user interface matters. These include colours of alarm lights or shapes of icons, most of which are not unique to AVs, and are well covered by standards and reference books. But human concerns extend to things such as:

- The roles of humans and computers - possible mismatches between expectations built into the AVs and real human behaviours, and vice versa
- Degrees of trust and dependence and their match/mismatch to trustworthiness/dependability
- People's ability to complement the automation and vice versa (e.g. dynamic transitions between automatic and manual modes)

The industry is aware of such issues, but evidence suggests that design approaches may not always fully accommodate them. Analyses of recent road accidents involving AVs [5] strongly suggest that failures that are typically attributed to human error may actually have their root in wrong design assumptions. In [5] the authors appeal for consideration of "human factors such as trust, complacency, and awareness" in the design of AVs. Human factors research on unintended detrimental effects of automated systems on computer-supported human performance has been conducted in a variety of domains [6], including AVs [7].

The human/social components

The list of human components when studying AVs is quite extensive. One obvious human component is the individual driver/passenger inside an AV, who can be examined in at least two roles:

- As someone who interacts with the AV in which they ride – the focus is on how well the human and the automated component cooperate, hand over control to each other (or wrestle control from the other), etc
- As someone who may or may not accept the automation – the focus is on whether they will buy an AV or activate the AV functions rather than always driving manually

It is important to also consider other human road users. Consideration must be given to whether they behave in ways the AV can live with, and/or predict reasonably well. In turn,

the behaviour of AVs must be made comprehensible so that the humans will behave sensibly around them. Consideration must also include potential anti-social behaviour by these road users.

Building AVs' models of the human and social components

When designing automated systems that are going to interact with human agents, it is necessary to understand and model the human behaviour, as the automated component will be expected to anticipate, support and appropriately respond to the people it interacts with.

Arguably, one the biggest challenges in the design of AVs is to accurately (and usefully) model the social context in which they will be deployed [9], with its complexity, heterogeneity and unpredictability ("weird stuff happens every day" [10]). This applies to both the behaviour of the individual driver in an AV [11] (see discussion of the individual driver/passenger below) and the interactions between AVs and pedestrians and other road users [12] (see discussion of other road users below).

Similarly, if the correct functioning or safety of the AV-human system requires a particular type of behaviour from a human operator in particular circumstances, it is crucial to determine whether the person will be able to exhibit the expected behaviour. For example, the Tesla autopilot requires that the human driver remains vigilant, monitoring the road and the automated displays, while the automated system is doing most of the driving. Is this an ability that can be reasonably expected from the average human driver? Does empirical research about human performance support this design assumption? This is relevant to adaptation (see the discussion of human adaptation to AVs below).

The individual driver/passenger

A lot of research has focused on the individual human driver in partly-automated vehicles (SAE Levels 1-3 [13]). Empirical studies of a human driver's behaviour during automated driving have been conducted for some time now [14][15][16]. One goal is to anticipate the effects of vehicle automation on the driver with a view to improved design strategies. This challenge can be viewed as replicating in AV design what psychologists term a "theory of mind" [17][18], that is, humans' ability to extrapolate from their own internal mental states to estimate how others might react. Humans are good at anticipating other people's intentions and reactions and, as a result, they can predict the likelihood of certain actions from other people based on (sometimes subtle) behavioural patterns.

Challenges regarding the modelling of the individual human driver include:

- Individual differences (e.g. different driving styles, personalities, cognitive strategies, attitudes to risk taking [19], etc.) [20]
- Cultural differences (e.g. driving styles and attitudes vary from country to country, even within the same country)
- The likelihood that AV models of driver behaviour will get outdated as the human agents get accustomed – and successively adapt – to automated capabilities [21]

Other road users

Modelling the social context goes beyond the individual riding the AV. It should also consider the heterogeneous road users, including human-driven cars and those known as "vulnerable road users" (VRUs) such as pedestrians and cyclists, and even animals. Particularly challenging is the prediction of intent by road users, which some identify as "the big problem with self-driving cars" [22] (e.g. how can an AV read a pedestrian's body language or predict whether a parked car will suddenly move into the lane). It is an issue that has not yet received as much attention as other socio-technical aspects of AVs. Recent research in the area includes heuristic models of pedestrian intention estimation to assist AV decisions when approaching pedestrians (e.g. designing an AV to predict whether an approaching pedestrian will cross the road or yield to the vehicle) [23].

Human take-over

Much research has been devoted to human take-over from automated features and the factors that influence human behaviours in those situations [24]. This has been studied predominantly in the context of "highly automated" or "conditional automated" driving (typically, level 3 in SAE's taxonomy [25]), where the automated system performs all or most of the driving without human supervision but will still require human intervention (take-over) in situations that cannot be handled by the automated system (e.g. reaching a system boundary due to sensor limitations, ambiguous environment observations, a sudden lead car braking, an obstacle suddenly appearing on the road). In those cases, the driver must be able to take over control within a reasonable amount of transition time. Human take-over may occur in response to an automated "take-over request" (TOR) or be self-initiated by the driver without automated prompting.

Aspects of human take-over in "highly automated driving" that have been in the literature include:

- The design of automated TORs (e.g. whether these should be auditory or visual) to promote human compliance and safety [26]
- "Time budget" (the time between the onset of an event and an impending crash) and human reaction time to TORs; the literature suggests that a safety buffer of 8-10 seconds is sufficient for a driver to take over the driving task comfortably [27]
- The impact on the safety of "non-driving related tasks" (or secondary tasks) that a human driver may be performing just before take-over (e.g. playing games, watching films, reading, texting, etc.) [28][29][30][31]
- Trust of automation, which seems to increase with drivers' experience of highly automated driving [32]
- Drivers' age, which affects drivers' trust of automation (e.g. older drivers rate vehicle automation more positively than younger ones) [32] and the way they respond to hazards and take over requests, although both older and younger drivers can solve critical traffic events and both adapt to the experience of take-over situations in the same way [33]

For a recent (at the time of writing) review of empirical and modelling research on take-overs in highly automated driving see [34].

Take-over can also be considered in higher levels of automation, for example, SAE's Level 4, where human drivers are not required for the AV's safe driving but may assume control after exiting the ODD. Recent research looks, for example, at human take-over after sleeping [35].

The challenge of incorporating what human factors research has found out about human drivers' behaviours and cognition into the design of intelligent software in AVs is not trivial. But there exists a body of evidence of how drivers perform and existing models of human driving (e.g. braking and steering) in manual conditions which are applicable to modelling how humans respond to TORs [35]. Such models and analyses are being used to develop assistance systems to predict driver's take-over behaviour (e.g. sideswipe manoeuvres) when the AV is handling the driving task and so automatically provide more useful take-over suggestions to the human driver [36].

Humans' perceptions and mental models of the AV

Another important issue is to ensure that there is proper coordination and communication between the automated driving components and the human participants.

There are at least two considerations:

- **Clear communication to the human driver** of the level of automation (or mode of operation) that an AV is working at in order to avoid mode confusion (e.g. cases where truck drivers used Level 2 autopilot features as though they were driving in a Level 3 mode, leading to lapses of attention [37]). Quoting [38], "an AV up to SAE Level 4 should inform its driver about the AV's capabilities and operational status, and ensure safety while changing between automated and manual modes". The key is to provide the right amount and kind of information about the AV's actions to the human driver (not too much or not too little) [39].
- **Clear communication to road users** of the "intentions" of the (fully) AV, via, for example, external Human-Machine Interfaces (HMIs) [41]. This area is receiving increasing attention from the community lately [41][42][43][44]. Solutions proposed include LED signs mounted on the outside of the vehicle announcing status such as "going now, don't cross" vs. "waiting for you to cross", as well as on-car projections to attached screens.

Human adaptation to AVs

Inevitably, the introduction of AVs will change people's behavioural patterns, whether it is the behaviours of the human driver of an AV [46], or of human drivers of non (or less) automated cars on the same road, or of VRUs.

People will adapt to AVs and anticipating the consequences of such adaptation is a challenge. Furthermore, many adaptation issues will evolve gradually if AVs are slowly introduced and developed, hence we will need methods for monitoring how the AVs evolve and how human adaptation follows.

Below are three examples of studied human (mal)adaptation to AVs:

- Risk compensation – a common consequence of the adoption of safety technologies is that it can lead to aggressive or reckless behaviours that would not take place without the safety features [47][48][49]
- Abuse of the conservative, cautious, behaviours of AVs, in particular by other road users. For example, a poll by LSE and Goodyear, together with City University, London, and other European universities, found that human drivers of non-automated vehicles on the road will "bully" AVs, which they perceive as prudent and abiding by the rules, so susceptible of being taken advantage of on the road [50]
- Over-trust of the AV's capabilities, leading to overreliance (see discussion of automation bias below)

There are other adaptation risks, which do not have much attention in the literature, for example:

- Situations where drivers shift between vehicles that are not standardised in their capabilities. For example, nowadays when someone rents a car, it may have: cruise control with automatic braking to keep a safe distance from the vehicle ahead; warning for obstacles when reversing etc. Suppose that someone owning a car with extensive assistive/autonomous features rents or borrows a different car or buys a new model. Recognising exactly what features are present in the latter may not be straightforward, but even after understanding the differences, a driver may still fall back into behaviours and practices learned on the habitual or previous car, in which they may, inappropriately, automatically delegate some tasks to the automation.
- Restricted environments where most of the vehicles are fully automated vehicles (SAE Level 4; e.g. trucks in a military environment) and staff have had the opportunity to adapt their mode of operation. It is likely that when vehicles with human drivers are incorporated into the restricted environment, theirs then become the unexpected (less predictable) way of driving, to which staff will need to re-adapt.

Below, we briefly discuss a special scenario of human adaptation to AVs: the case of safety drivers in road testing [51].

A special case scenario: the safety driver

Adaptation can also affect the role of the safety assessment in road testing [52][51]. In so far as the safety driver is the last layer of defence against possible errors of the AV (including its built-in safety layers), the concern with adaptation is that the safety drivers may become progressively less effective in this role, for at least two reasons:

1. If the AVs are acceptably safe, the drivers risk becoming accustomed to their intervention being very seldom needed. This can impair their ability to:
   - maintain proper vigilance – humans are poorly equipped for continuous monitoring tasks
   - recognise events that require intervention – humans, consciously or not, recognise that these events are rare and so indications of hazardous situations may well be false alarms
   - properly intervene when needed – due to losing situation awareness
2. If safety drivers' interventions are used to detect unforeseen hazardous behaviour situations these can be corrected in the AV. Because the hazardous behaviours that they detect more easily will have been reported and corrected, the AVs' hazardous

behaviours over time will tend to be those that are less feasible for the safety drivers to detect and deal with

Automation bias

"Automation bias" is a well-documented phenomenon, which refers to situations where human operators perform worse when using technology that is designed to support them in a particular task than when they perform the same task unsupported by the technology. Automation bias was originally studied in monitoring tasks in aviation [53] but has now been documented in many other automated domains, including AVs [54][55]. An example of automation bias on the road would be a driver who misses an obstacle in front of the vehicle because the automated system fails to alert the human to the presence of the obstacle. Had the driver been using a non-automated car, they would not have missed it. In [54], the authors describe accidents involving AVs as "uncanny" accidents that "any reasonably attentive, sober driver would easily avoid".

These unintended and undesirable situations are commonly explained as the result of the human becoming less vigilant because of complacency which may be induced by over-trust in the automated system leading to over-reliance [39]. Maladaptation is also attributed to people's failure to properly follow the automated system's instructions [31]. In a case study which looked at automation bias in health informatics [2] it was found that operators would still be prone to automation-induced errors even if they remained vigilant, appropriately followed the instructions of use, and reported that they did not trust the technology.

An important issue to consider is the level of engagement of human operators with the driving task in AVs [50]. Driving automation, as is currently implemented, isolates drivers from functional experience – it may even take away from them some of the pleasures people often associate with driving. Given the black box nature of many existing self-driving features, drivers may not have a clear understanding of the automated capabilities. As a result, they will end up creating a mental model of said capabilities, generalising after a few interactions, and perhaps infer that the automated system can do more than it actually can. This mismatch, combined with reduced engagement with the driving experience, may lead to reduced vigilance and reduced situation awareness, sometimes wrongly expecting that the automated system will be able to handle hazards for which it has not been designed.

References

- [1] Brown, B., & Laurier, E. (2017, May). The trouble with autopilots: assisted and autonomous driving on the social road. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (pp. 416-429). ACM.
- [2] Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019, January). Fairness and abstraction in sociotechnical systems. In Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 59-68). ACM.
- [3] SaFAD (2019) "Safety First for Automated Driving". White paper.
- [4] Sadigh, D., Sastry, S., Seshia, S. A., & Dragan, A. D. (2016, June). Planning for autonomous cars that leverage effects on human actions. In: Robotics: Science and Systems (Vol. 2).
- [5] Clancy, J. & Jarrahi, M. H. (2019). Breakdowns in Human-AI Partnership: Revelatory Cases of Automation Bias in Autonomous Vehicle. Preprint.

https://www.researchgate.net/publication/337910493_Breakdowns_in_Human-AI_Partnership_Revelatory_Cases_of_Automation_Bias_in_Autonomous_Vehicle

- [6] Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human factors, 39(2), 230-253.
- [7] Stanton, N. A., & Young, M. S. (2000). A proposed psychological model of driving automation. Theoretical Issues in Ergonomics Science, 1(4), 315-331.
- [9] Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019, January). Fairness and abstraction in sociotechnical systems. In Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 59-68). ACM.
- [10] Monroe, D. (2019). I don't understand my car. Communications of the ACM, 62(8), 18-19
- [11] Merat, N., & Jamson, A. H. (2009). How do drivers behave in a highly automated car? 2009 Driving Assessment Conference.
- [12] Rasouli, A., & Tsotsos, J. K. (2019). Autonomous vehicles that interact with pedestrians: A survey of theory and practice. IEEE Transactions on Intelligent Transportation Systems.
- [13] SAE J 3016-2018 taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles, [online] Retrieved, December 4, 2019, from: https://webstore.ansi.org/Standards/SAE/SAE30162018
- [14] 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.
- [15] Stanton, N. A., & Young, M. S. (2000). A proposed psychological model of driving automation. Theoretical Issues in Ergonomics Science, 1(4), 315-331.
- [16] Merat, N., & Jamson, A. H. (2009). How do drivers behave in a highly automated car? 2009 Driving Assessment Conference.
- [17] Monroe, D. (2019). I don't understand my car. Communications of the ACM, 62(8), 18-19
- [18] Surden, H., & Williams, M. A. (2016). Technological opacity, predictability, and self-driving cars. Cardozo L. Rev., 38, 121.
- [19] Zeeb, K., Buchner, A., & Schrauf, M. (2015). What determines the take-over time? An integrated model approach of driver take-over after automated driving. Accident Analysis & Prevention, 78, 212-221.
- [20] Shutko, J., Osafo-Yeboah, B., Rockwell, C., & Palmer, M. (2018). Driver Behavior While Operating Partially Automated Systems- Tesla Autopilot Case Study (No. 2018-01-0497). SAE Technical Paper.
- [21] Grane, C. (2018). Assessment selection in human-automation interaction studies: The Failure-GAM2E and review of assessment methods for highly automated driving. Applied ergonomics, 66, 182-192.
- [22] Brooks, R. (2017). The big problem with self-driving cars is people and we'll go out of our way to make the problem worse. [online] Retrieved, December 6, 2019, from: http://acl.kaist.ac.kr/techtrend?mod=document&uid=302.
- [23] Camara, F., Merat, N., & Fox, C. W. (2019, October). A heuristic model for pedestrian intention estimation. In 2019 IEEE Intelligent Transportation Systems Conference (ITSC), pp. 3708-3713.
- [24] McDonald, A. D., Alambeigi, H., Engström, J., Markkula, G., Vogelpohl, T., Dunne, J., & Yuma, N. (2019). Toward computational simulations of behavior during

automated driving takeovers: a review of the empirical and modeling literatures. Human factors, 61(4), 642-688.

- [25] SAE J 3016-2018 taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles, [online] Retrieved, December 4, 2019, from: https://webstore.ansi.org/Standards/SAE/SAE30162018
- [26] Roche, F., Somieski, A., & Brandenburg, S. (2018). Behavioral changes to repeated takeovers in highly automated driving: effects of the takeover-request design and the nondriving-related task modality. Human factors, 0018720818814963.
- [27] Melcher, V., Rauh, S., Diederichs, F., Widlroither, H., & Bauer, W. (2015). Take-over requests for automated driving. Procedia Manufacturing, 3, 2867-2873.
- [28] Dogan, E., Honnêt, V., Masfrand, S., & Guillaume, A. (2019). Effects of non-driving-related tasks on takeover performance in different takeover situations in conditionally automated driving. Transportation research part F: traffic psychology and behaviour, 62, 494-504.
- [29] Naujoks, F., Purucker, C., Wiedemann, K., & Marberger, C. (2019). Noncritical State Transitions During Conditionally Automated Driving on German Freeways: Effects of Non–Driving Related Tasks on Takeover Time and Takeover Quality. Human factors, 61(4), 596-613.
- [30] Roche, F., Somieski, A., & Brandenburg, S. (2018). Behavioral changes to repeated takeovers in highly automated driving: effects of the takeover-request design and the non-driving related task modality. Human factors, 0018720818814963.
- [31] Wandtner, B. (2018). Non-driving related tasks in highly automated driving: Effects of task characteristics and drivers' self-regulation on take-over performance.
- [32] Dunn, N., Dingus, T., & Soccolich, S. (2019). Understanding the Impact of Technology: Do Advanced Driver Assistance and Semi-Automated Vehicle Systems Lead to Improper Driving Behavior? AAA Foundation for Traffic Safety.
- [33] Körber, M., Gold, C., Lechner, D., & Bengler, K. (2016). The influence of age on the take-over of vehicle control in highly automated driving. Transportation research part F: traffic psychology and behaviour, 39, 19-32.
- [34] McDonald, A. D., Alambeigi, H., Engström, J., Markkula, G., Vogelpohl, T., Dunne, J., & Yuma, N. (2019). Toward computational simulations of behavior during automated driving takeovers: a review of the empirical and modeling literatures. Human factors, 61(4), 642-688.
- [35] Hirsch, M., Diederichs, F., Widlroither, H., Graf, R., & Bischoff, S. (2019). Sleep and take-over in automated driving. International Journal of Transportation Science and Technology.
- [36] Lotzetal (2019). An adaptive assistance system for subjective critical driving situations- understanding the relationship between subjective and objective complexity. Proceedings of HFES-Europe 2019.
- [37] Bieg, H-J., Daniilidou, C., Michel, B., Sprun, A.(2019). Task load of professional drivers during level 2 and 3 automated driving. In Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2019 Annual Conference. ISSN 2333-4959.
- [38] Kyriakidis, M., de Winter, J. C., Stanton, N., Bellet, T., van Arem, B., Brookhuis, K., ... & Reed, N. (2019). A human factors perspective on automated driving. Theoretical Issues in Ergonomics Science, 20(3), 223-249.

- [39] Koo, J., Kwac, J., Ju, W., Steinert, M., Leifer, L., & Nass, C. (2015). Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. International Journal on Interactive Design and Manufacturing (IJIDeM), 9(4), 269-275.
- [41] Lee (2019) Understanding the Messages Conveyed by Automated Vehicles AutomotiveUI '19, September 21–25, 2019, Utrecht, Netherlands.
- [42] Nguyen, T. T., Holländer, K., Hoggenmueller, M., Parker, C., & Tomitsch, M. (2019, September). Designing for Projection-based Communication between Autonomous Vehicles and Pedestrians. In Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 284-294). ACM.
- [43] Surden, H., & Williams, M. A. (2016). Technological opacity, predictability, and self-driving cars. Cardozo L. Rev., 38, 121.
- [44] Mahadevan, K., Somanath, S., & Sharlin, E. (2018, April). Communicating awareness and intent in autonomous vehicle-pedestrian interaction. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (p. 429). ACM.
- [46] Grane, C. (2018). Assessment selection in human-automation interaction studies: The Failure-GAM2E and review of assessment methods for highly automated driving. Applied ergonomics, 66, 182-192.
- [47] Brown, B., & Laurier, E. (2017, May). The trouble with autopilots: assisted and autonomous driving on the social road. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (pp. 416-429). ACM.
- [48] Yamada, K., & Kuchar, J. K. (2006). Preliminary study of behavioral and safety effects of driver dependence on a warning system in a driving simulator. IEEE transactions on systems, man, and cybernetics-Part A: Systems and humans, 36(3), 602-610.
- [49] Millard-Ball, A. (2016). "Pedestrians, Autonomous Vehicles, and Cities," Journal of Planning Education and Research, pp. 1-7 (DOI: 10.1177/0739456X16675674); at https://bit.ly/2hhYrxV.
- [50] LSE (2016). "Autonomous Vehicles - Negotiating a Place on the Road. A study on how drivers feel about Interacting with Autonomous Vehicles on the road. EXECUTIVE SUMMARY.
- [51] Zhao, X., Robu, V., Flynn, D., Salako, K., & Strigini, L. (2019). Assessing the safety and reliability of autonomous vehicles from road testing. In the 30th Int. Symp. on Software Reliability Engineering (ISSRE), IEEE, Berlin, Germany, 2019, in press.
- [52] Koopman, P., & Osyk, B. (2019). Safety Argument Considerations for Public Road Testing of Autonomous Vehicles (No. 2019-01-0123). SAE Technical Paper.
- [53] Skitka, L.J., Mosier, K., Burdick, M.D. (1999). Does automation bias decision making? International Journal of Human-Computer Studies, 51 (5), 991-1006.
- [54] Clancy, J. & Jarrahi, M. H. (2019). Breakdowns in Human-AI Partnership: Revelatory Cases of Automation Bias in Autonomous Vehicle. Preprint. https://www.researchgate.net/publication/337910493_Breakdowns_in_Human-AI_Partnership_Revelatory_Cases_of_Automation_Bias_in_Autonomous_Vehicle
- [55] Dunn, N., Dingus, T., & Soccolich, S. (2019). Understanding the Impact of Technology: Do Advanced Driver Assistance and Semi-Automated Vehicle Systems Lead to Improper Driving Behavior? AAA Foundation for Traffic Safety.