Misplaced Trust?

Michael Fisher, Nick Reed, Joseph Savirimuthu

Introduction

Driverless cars are on the way. Most major motor manufacturers are developing them, and all expect them to be on our roads sooner, rather than later [1]. When we drive, we believe we are in control. Yet, technologies such as cruise control, lane control and, soon, "platooning" or "vehicle convoying" are beginning to take away our direct control of vehicle speed/direction. While we may still make the key decisions—to overtake, to stop, etc—intelligent technologies increasingly control the basic "path following" activities of our vehicle. Furthermore, the UK Department for Transport [2] states:

> *"By 2040, experts expect a world of connected vehicles and road users, where `semi-autonomous' and `autonomous' control of vehicles will be part of life.... Innovative ways to make vehicles cooperate with one another, such as the `platooning' approach for heavy vehicles on strategic roads, have the potential to make our roads work better for everyone."*

While driverless cars are expected to unleash a revolution in the transport sector, the considerable investment, innovation and research that could follow will be hindered if ongoing legal barriers, ethical uncertainties and consumer safety concerns are not resolved. Specifically, user *trust* is a key, and highly subjective, element without which driverless cars will fail to have mass appeal.

In this article we argue that, although the basic technologies for producing driverless cars have been developed, three core areas have been neglected: human factors; legal aspects; and software verification. We assert that all three will eventually have to be tackled but they should be central to the design and development of these vehicles *now*, rather than as an afterthought. Furthermore, we believe that all three impact upon subjective notions of trust and, without advances in these areas, the public's trust in driverless cars will remain lacking. While this article addresses *semi-autonomous* vehicles, where a human is involved, the issues we tackle are even more important as we move towards truly autonomous vehicles.

Driverless Cars

Present day electronic stability control (ESC) systems have reduced the number and severity of crashes [6]. If sensors detect that a vehicle is slipping, ESC seamlessly applies the braking actuators for individual wheels to help a driver retain control of the vehicle. Responsibility for control of the vehicle is shared and these systems are seen as supporting the driver’s intention. As autonomous technologies develop, the balance between driver and vehicle in their responsibility for safe control of the vehicle will shift towards the vehicle. Implemented correctly, as with ESC, this can be expected to bring a reduction in collision risk. However, where ESC affects vehicle control in more extreme situations, more generalised automation may have a greater influence on how a driver approaches the task of vehicle control and navigation. Such issues have been studied in the context of aviation where autopilot systems are well understood. However, driving is undertaken in a disordered environment where abrupt changes in situation are more common and where operators have experienced a less rigorous training regime. Where software is responsible for safety critical decisions and actions but responsibility for safe control of the vehicle is shared, the driver needs to know if the system is active and working correctly, what the system is likely to do next (and why) and when the system may disengage (and be sufficiently aware of the driving situation in order to resume control of the vehicle safely). There are clear interface and feedback design challenges in ensuring the driver is kept adequately *in-the-loop* with respect to the status of the vehicle systems and traffic situation.

Whilst the early implementations of autonomy require that the driver to maintain alertness and attention to the driving scene, already demonstrators have shown that autonomous systems that allow the driver to engage in other tasks are technically feasible. For example, SARTRE [10] was just one example project that has demonstrated electronic coupling of vehicles on a highway such that participating drivers following the lead vehicle need pay no attention to the driving task and can read, eat or send text messages whilst the road train is operating. After safety, recapturing the time spent driving and applying it to more useful tasks is one of the key motivations behind the application of autonomy in road transport. However, motion sickness might limit the time that can realistically be reclaimed. It is a commonly experienced phenomenon that passengers are more susceptible to such discomfort than drivers. This risk is heightened for passengers who choose to read or watch/interact with a small screen (tablet or smartphone) so that they are not observing the external environment. While autonomous systems may be able to drive the vehicle with optimal smoothness, the utility of time spent being driven rather than driving may be limited by our own human frailties.

These new technologies clearly present difficulties. Yet they only scratch the surface of autonomy. By `autonomy' we mean the vehicle's ability to make its own decisions about what to do and when to do it. While many of the above technologies simply adapt to their environment, responding automatically to environmental changes, there is very much more that autonomous vehicles can do. Indeed, fully autonomous vehicles are on the way as well, though not just yet (while there are several proposals for more advanced vehicles, these are either employed on restricted routes or show quite limited autonomy) [2]:

> *"Fully autonomous cars remain a further step, and for the time being drivers will have the option (and responsibility) of taking control of the vehicle themselves. Vehicle manufacturers and their systems suppliers continue to explore the opportunities for full autonomy. Further progress will depend foremost on ensuring public safety and on updating the law to take account of the new technology."*

Once we move to fully autonomous vehicles, then it will be the controlling software that makes all the decisions that the human driver used to make: when to overtake; when to turn off the motorway; how to react to unexpected situations; etc. As we move towards this stage, then the above concerns about "legality", "safety" and "trust" will become even more important.

Human Factors

Human factors is the scientific discipline that seeks to understand and optimise how people use and interact with systems, combining principles of applied psychology, design and engineering. The subject developed rapidly through studies of pilot interaction with the controls and displays in combat aircraft from World War II onwards. These transformed consideration of the human operator, with a discrete set of capabilities and limitations, as but one component in overall system performance.

Roads and vehicles are key elements in delivering personal mobility and the movement of goods; the performance of the human operators of these vehicles has provided fertile ground for research into how human factors can improve road safety. The need to understand the influence of drivers in collision causation was brought into sharp relief by the finding that human error was (and remains) a contributory factor in 95% of accidents [12].

Minimising driver error has therefore provided impetus for much of the human factors research in road safety with studies investigating issues such as hazard perception (see [16]), training interventions (e.g. [17]), fatigue (e.g. [15]) and distraction (e.g. [13]). Driver assistance systems such as anti-lock brakes and electronic stability programs have contributed to a reduction in the frequency and severity of collisions. However, advances in computing power, communications, image/data processing and sensor technology have led to the genuine prospect that onboard systems may be able to take some- and eventually all of the responsibility for safe control of the vehicle. This progression has been captured in the classifications of vehicle automation proposed by BASt [14], NHTSA [18] and the SAE [19].

Whilst the technical challenges to vehicle automation are being steadily overcome, the behavioural changes that the introduction of vehicles with higher levels of automation may trigger are a cause for concern. A considerable number of drivers engage with other technologies [ref] such that their attention is drawn away from the activities required for safe driving – the definition of driver distraction [20]. In partially and highly automated vehicles, where the requirement for engagement in the driving task is reduced, the temptation will be even greater. However, in such vehicles, there will be occasions when the driver is required to resume control. Managing automation mode transitions when a driver may be distracted poses numerous questions. In switching to an automated mode, how and when does the vehicle communicate to the driver the tasks for which the system is now responsible? To what extent is the driver monitored to ensure that they are sufficiently engaged with the driving task when the vehicle has control (Eye tracking? One hand on steering wheel?)? How long does the distracted or sleeping driver need to achieve sufficient awareness of the driving situation such that they can safely re-engage with the driving task? What information and cueing mechanisms will be most effective in managing this process? How does the vehicle manage if the driver is unable or refuses to resume control? In returning control to the driver, does the vehicle always return to full manual control (no automation) or does the vehicle step down through automation levels gradually? While engineers deliver technical solutions to enable automated driving, the answers to each of the questions may be critical in ensuring that drivers' experience of automated vehicles is safe and enjoyable.

In fully automated vehicles, driver distraction is not an issue as the vehicle takes full responsibility for safe control of the vehicle between origin and destination. This opens up time for the traveller that could be put to good use. Estimates about the productivity benefit that this yields are impressive (see [21]). However, the experience of being passively driven and trying to read text or interact with a screen accentuates facets of road travel that promote motion sickness [ref]. Those claims about time spent in an automated vehicle being productive should be validated in this context as a significant number of people may find trying to work as a passenger in the car too uncomfortable,

A further claim is that automated vehicles give the promise that independent mobility may be achievable for those who at present find it difficult or are unable to drive due to medical conditions [ref]. Whilst this is certainly true, care must be taken in predicting the extent to which such travellers will be able to have their needs met by such vehicles. For many, door-to-door navigation is the only option and the autonomous vehicle must be able to achieve this goal – arriving at a point even as little as 100m from the destination may present a more challenging situation than having never started the journey in the first place.

In the language of robotics, road transport is a 'known' but 'unpredictable' environment. 'Known' in that a system can be programmed with the formal rules of the road and provided with a complete map of the driving environment for navigation; but it is 'unpredictable' in that it can never be determined what combination of traffic, weather or road conditions one might encounter on any given journey. Consequently, the tiniest coding error or oversight that may not be apparent in a rigorous testing regime may become a critical contributory factor when circumstances align resulting in a collision. It should therefore be noted that although relying on systems for vehicle control may reduce the risk of driver error, it does not remove all opportunities for human error to cause collisions. The determination of liability in such circumstances is one of but a number of legal issues that may be of critical importance to the eventual deployment of automated vehicles.

Legal Aspects

The driver has long been the focal point of transport industry and policymaking. Under the *Vienna Convention*, "[e]very driver shall at all times be able to control his vehicle…" The 1968 Vienna Convention on Road Traffic also allocates responsibility in the driver of the vehicles. Drivers are regarded as having particular attributes. For example, Article 8 (2) stipulates that drivers possess physical and mental abilities. These attributes incorporate knowledge and skill to enable that the vehicle is driven in a way that does not expose the driver or others to danger or harm. Within the UK, both the *Road Traffic Act 1988 (as amended) and* the *Highway Code* for example, reinforce the view of a human driver being in control. The *Road Traffic Act 1988*, the *Road Vehicles (Construction and Use) Regulations 1986* and the *Highway Code* already extend to advanced motoring technologies that assist drivers in operating, steering and navigating vehicles. Many motor vehicles come equipped as standard that assist drivers, such as Anti-lock Braking Systems (ABS), cruise and adaptive speed control and parking sensors. ABS, it will be recalled, does not require human instructions or direct control. Other technological innovations such as Advanced Emergency Braking Systems (AEBS) or Electronic Stability Control (ESC) enable the motor vehicle to operate in a safe manner without prior human driver intervention. These technologies involve sophisticated software to process information and direct the driving process. Autonomous and semi-autonomous systems will lead to fundamental reassessment of how best to reflect human-machine collaboration in regulatory systems. The early indications are that we may not necessarily have to design new laws as such. However, as semi-autonomous technologies are increasingly integrated into motor vehicles, the existing regulatory framework will need to reflect the fact human drivers may not be in full control of the vehicle. With regard to issues of safe operation of semi-autonomous motor vehicles or liability arising from collisions with other vehicles, the issue of fault have to be addressed. As a general rule, as is the case under the current road traffic regulations, the person using the car will be regarded as liable. However, some amendments may be needed

A better way of approaching liability issues would be to ascertain if the collision is due to a product defect or the negligence of the human driver. Sometimes, it could be due to both – for example, it may be reasonable to expect a driver to take due care when parking rather than solely rely on parking sensors.

Product Liability

Semi-autonomous cars are not immune from software or design faults. In the event collisions or accidents caused by product defects, recourse under tort law is available (Donoghue v Stevenson [1932] A.C. 562). Product liability under tort law extends beyond persons who have a contractual relationship with the purchaser of the motor vehicle, and will include manufacturers of software and those responsible for certifying the quality and safety the vehicle. The Consumer Protection Act 1987, as amended by Council Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products), establishes a strict liability regime. The significance of a strict liability regime is that purchasers of autonomous vehicles will not have to prove negligence in holding manufacturers or retailers accountable for defective products [11]. For example, if an autonomous motor vehicle unexpectedly speeded up and crashed into a stationary car due to malfunctioning software, the manufacturer would be held liable under the Consumer Protection Act 1987. Section 2 of the Consumer Protection Act 1987 provides that where "any damage is caused wholly or partly by a defect in a product", liability for the damage will extend to:

"*(2) —*

(a) the producer of the product;
(b) any person who, by putting his name on the product or using a trade mark or other distinguishing mark in relation to the product, has held himself out to be the producer of the product;
(c) any person who has imported the product into a member State from a place outside the member States in order, in the course of any business of his, to supply it to another.

(3) Subject as aforesaid, where any damage is caused wholly or partly by a defect in a product, any person who supplied the product (whether to the person who suffered the damage, to the producer of any product in which the product in question is comprised or to any other person) shall be liable for the damage if—

(a) the person who suffered the damage requests the supplier to identify one or more of the persons (whether still in existence or not) to whom subsection (2) above applies in relation to the product;
(b) that request is made within a reasonable period after the damage occurs and at a time when it is not reasonably practicable for the person making the request to identify all those persons; and
(c) the supplier fails, within a reasonable period after receiving the request, either to comply with the request or to identify the person who supplied the product to him."

If the evidence however shows that the damage was caused by human driver intervention rather than manufacturing, design or warning defects in the product, liability under the Consumer Protection Act 1987 will not arise (McGlinchey v General Motors UK Ltd [2011] CSOH 206). Section 3 of the Consumer Protection Act 1987 incorporates a "consumer expectations" test when determining what constitutes a "defective product". The section defines a "defect" as present if "the safety of the product is not such as persons generally are entitled to expect; and for those purposes "safety", in relation to a product, shall include safety with respect to products comprised in that product and safety in the context of risks of damage to property, as well as in the context of risks of death or personal injury.

In determining, for the purposes of subsection (1) above, what persons generally are entitled to expect in relation to a product all the circumstances shall be taken into account, including—

(a) the manner in which, and purposes for which, the product has been marketed, its get-up, the use of any mark in relation to the product and any instructions for, or warnings with respect to, doing or refraining from doing anything with or in relation to the product;
(b) what might reasonably be expected to be done with or in relation to the product; and
(c) the time when the product was supplied by its producer to another;

and nothing in this section shall require a defect to be inferred from the fact alone that the safety of a product which is supplied after that time is greater than the safety of the product in question."

To ensure that there are no misaligned expectations in respect of the autonomous motor vehicle, it is imperative that distributors, manufacturers and retailers comply fully with the General Product Safety Regulations 2005 (as amended by Directive 2001/95/EC).

Law of Negligence

Assuming that human error, rather than product defect, is the cause of the collision, under the tort of negligence, the driver will be deemed to be responsible. Under the tort law, drivers are deemed to owe a duty of care to passengers and other road users. The standard of care is the same regardless of whether a driver uses a traditional or a highly autonomous motor vehicle (Nettleship v Weston [1971] 2 Q.B. 691). Given the infancy of autonomous vehicles, it is very likely that the default rule would be that human drivers will be expected to demonstrate reasonable levels of competence (Mansfield v Weetabix Ltd [1998] 1 W.L.R. 1263). For example, drivers will be expected to exercise due care and skill when features such as automatic parking, adaptive cruise control and lane keeping are engaged. If the human driver fails to monitor the vehicle or creates a foreseeable risk of damage or harm, that would be prima facie evidence of a breach of duty of care (Goad v Butcher [2011] EWCA Civ 158).

Formal Software Verification

Although the legal framework for increasingly autonomous vehicles is not yet in place, researchers have begun to tackle the routes to ensuring the legality and safety of such vehicles. The authorities responsible for certification may be able to lay the ground rules, but this is not as straightforward as it might first appear. Although extensive testing can help, it is likely that we will have to turn to more comprehensive techniques for “proving” properties of the new internal software such as *formal verification* used in critical systems [5]. This popular strand of research in Computer Science is concerned with the deep, Mathematical analysis of software and, in particular, providing logical justification that software will always match its formal requirements. Formal verification is particularly useful for analysing and evaluating software that plays a roles in human safety (i.e. *safety-critical*), such as software within power-stations, life-support systems or transportation systems.

In the case of driverless cars, key software must make the decisions that a human driver once made. Such decisions can range from the somewhat mundane, such as whether to turn the headlights on, to much more serious, such as whether to brake violently. These vehicles, and autonomous systems in general, are essentially controlled by software and, by isolating the core decision-making component, we have a chance to apply formal verification. Once we isolate the software that makes all the high-level decisions (those decisions that a human would have made) then we can explore the detailed working of these programs. (Note that this is not something you can easily, or possibly ever could, do with a human brain.) By using new formal verification techniques specifically developed for such autonomous decision-making [5], we can prove properties of the software making the high-level decisions within our autonomous systems. For example, we might be able to prove that the software controlling our system will never make bad decisions, or make decisions it believes to be dangerous.

Once we have such techniques for deep analysis of the autonomous decision-making core within our driverless car's software, the problem remains of *what* exactly to give as formal requirements. In the case of unmanned air vehicles, we have shown how such verification can be used to establish that the autonomous system's decision-making matches (at a basic level) the pilot's. Once we recognise that the key difference between an autonomous vehicle and the same vehicle controlled by a human driver/pilot is that the human is replaced by autonomous decision-making software, then we can attempt to prove that the software will *always* make the same decisions s a human pilot should. For example, in [7] we formally verified that the core software agent controlling and unmanned air vehicle made the decisions expected from the "Rules of the Air", i.e. the rules that pilots are required to learn and understand. For driverless cars we might do something similar, but need a clear description of the expected behaviour of human drivers. Once we have this description, we can formally verify that the core software will always make appropriate decisions.

Issues of Trust

Innovations such as driverless cars also pose an ethical-psychological dilemma. What are the ethical consequences of allowing human drivers to cede control or blaming software malfunction for a road fatality? As the driver increasingly becomes the passenger, what does this experience feel like [3]? Regardless of technological and legal developments, do we trust cars with these technologies enough to want to use them? These social and psychological aspects could well be the biggest barriers to adoption. In short, is society ready for truly driverless cars? This lack of trust is likely to be the biggest barrier to widespread take-up. However, trust issues are also being faced by other varieties of autonomous systems [8,9], and advances in those areas may well impact upon autonomy in the automotive sector.

Our assertion here is that public *trust* is a key issue, and that there are numerous factors that might affect this. We believe that the three aspects of human factors, legal issues, and software verification all impact upon trust. Do the public think such vehicles are safe? Do they feel comfortable using them? Do the public feel confident that the system will always make the best (safest, and for the passenger's benefit) decisions? These questions are crucial and, since all three will eventually have to be tackled, we assert that they should be central to the design and development of driverless cars *now*, rather than as an afterthought.

Finally, though it seems the realm of Science Fiction, there may well be ethical questions raised. Even before we get to fully autonomous vehicles, there are decisions the driverless car must make that have strong implications. Imagine an automated vehicle platoon on the highway. A human driver of a vehicle has ceded control of their vehicle to the platooning systems. If the driver chooses to take back control suddenly and turn the steering wheel sharply, should the system relinquish control back to the human driver if it is apparent that this manoeuvre will likely cause a crash? Should it follow the law and obey the human or should it invoke ethical principles and seek to protect human life? The answers to such questions may be critical to the successful deployment and uptake of automated vehicles.

References:

1. "Autos on autopilot: the evolution of the driverless car". Jon Excell, The Engineer. http://www.theengineer.co.uk/in-depth/the-big-story/autos-on-autopilot-the-evolution-of-the-driverless-car/1016859.article
2. "Action for Roads". Department for Transport, UK Government. https://www.gov.uk/government/publications/action-for-roads-a-network-for-the-21st-century
3. "Geneva Motor Show: Passengers' view of driverless cars". Russell Hotten, BBC News. http://www.bbc.co.uk/news/business-26433493
4. Vienna Convention, 1968.
5. "Verifying Autonomous Systems". Michael Fisher, Louise Dennis, and Matt Webster. *Communications of the ACM 56(9):84-93*, 2013. http://doi.acm.org/10.1145/2494558
6. “The Effectiveness of Electronic Stability Control in Reducing Real-World Crashes: A Literature Review” Susan A. Ferguson (2007), *Traffic Injury Prevention*, 8:4, 329-338, DOI: 10.1080/15389580701588949
7. "Generating Certification Evidence for Autonomous Unmanned Aircraft Using Model Checking and Simulation". Matt Webster, Neil Cameron, Mike Jump, and Michael Fisher. *Journal of Aerospace Information Systems 11(5):258-279*, May 2014.
8. "If you want to trust a robot, look at how it makes decisions". Michael Fisher. *The Conversation*, March 2014. http://theconversation.com/if-you-want-to-trust-a-robot-look-at-how-it-makes-decisions-24134
9. "Trustworthy Robotic Assistants" project. http://www.robosafe.org
10. “Safe Road Trains for the Environment” project. http://www.sartre-project.eu
11. Product Liability and Safety Law. https://www.gov.uk/product-liability-and-safety-law
12. “Factors Contributing to Road Accidents”. B. Sabey. TRRL LF976 Crowthorne: Transport Research Laboratory, 1983.
13. “The Effect of Text Messaging on Driver Behaviour”. N. Reed and R. Robbins. TRL PPR 387. Crowthorne: Transport Research Laboratory, 2008.
14. “BASt-study: Definitions of automation and legal issues in Germany”. T. M. Gasser and D. Westhoff. Presented at the *TRB Road Vehicle Automation Workshop*, Irvine, CA, 2012.
15. “Monotony of Road Environment and Driver Fatigue: A simulator study”. P. Thiffault and J. Bergeron. *Accident Analysis and Prevention 35(3): 381-391,* 2003.
16. “How can we produce safer new drivers? A review of the effects of experience, training, and limiting exposure on the collision risk of new drivers”. S. Helman, G. Grayson, and A. M. Parkes. TRL Insight Report (INS005). Crowthorne: Transport Research Laboratory, 2010.
17. “From research to commercial fuel efficiency training for truck drivers using TruckSim”. N. Reed, S. Cynk, and A. M. Parkes. In *Dorn, L. (Ed.), Driver behaviour and training*, volume IV (pp. 257-268). Aldershot, UK: Ashgate, 2010.
18. National Highway Traffic Safety Administration (NHTSA). Preliminary statement of policy concerning automated vehicles, 2013. Retrieved from http://www.nhtsa.gov/staticfiles/rulemaking/pdf/Automated_Vehicles_Policy.pdf
19. On-Road Automated Vehicle Standards Committee. Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems, 2014. Retrieved from http://standards.sae.org/j3016_201401
20. US EU Bilateral ITS Technical Task Force. Expert Focus Group on Driver Distraction: Definition and Research Needs. Berlin, Germany, 2010. Retrieved from http://ec.europa.eu/information_society/newsroom/cf/document.cfm?doc_id=1030
21. “Autonomous Cars: Self-Driving the New Auto Industry Paradigm”. Morgan Stanley *Blue Paper*, 2013.