Considerations Toward Driving Acceptance & Trust of Autonomous Vehicles

Connor Goddard

Department of Computer Science, Aberystwyth University
Aberystwyth, Ceredigion, SY23 3DB
Email: clg11@aber.ac.uk

***Abstract*—As recent technical accomplishments continue to accelerate the notion of driverless vehicles from distant possibility toward tangible reality, increasing attention within this domain is focussed on addressing the social, legal and philosophical issues that are expected to represent significant barriers in the path toward mainstream adoption of fully-autonomous vehicles within our existing societal frameworks.**

From surveying current literature obtained from a range of academic, industry and government sources, this study focusses on the emergence of four major factors found to significantly influence the considerations held by consumers toward the adoption of driverless technologies: driver liability, occupant safety, occupant control and understanding of their autonomous vehicles and ethical concerns arising from increased autonomy.

I. Introduction

In the year 1886, Karl Benz designed, built and patented the world's first petrol-powered automobile [**?**]. Now widely considered as the birth of the modern motor vehicle, this event would set alight a revolution in human transportation, becoming the driving force behind some of the most important technological innovations of the past two centuries.

As technology has moved into the digital age, the automotive industry has remained a major ambassador for new development. Satellite navigation, cruise control and automatic emergency braking all represent examples of autonomous technologies that have since become mainstream as part of our modern-day society.

With progress in automation and artificial intelligence showing no signs of slowing down, an increasing level of attention is being directed toward the discussion surrounding the notion of fully-autonomous vehicles, in which the human driver no longer has to maintain focus on the road ahead.

Over the past couple of years, discussion has poured out from the confines of industry and into the public sphere. This has been popularised by the launch of high profile projects such as the Google self-driving car [**?**], and the introduction of semi-autonomous driving technologies such as the Tesla Autopilot system [**?**].

Today, there is little doubt as to the expectation for autonomous vehicles to arrive on future roads. In a recent survey looking at public opinion toward automated driving, Kyriakidis et al. found 69 percent of 5000 public respondents predicted driverless cars to hold 50 percent of the total market share by 2050 [**?**]. Industry experts have exhibited an even greater level of optimism, believing that such technology will have become part of everyday life by as early as 2040 [**?**].

Upon face value, autonomous vehicles look set to provide significant safety, economic and environmental benefits. At the top of these advancements, we find promises of fewer traffic-related deaths and injuries; greater fuel efficiency; eased road congestion; and greater access to mobility for disadvantaged groups (including disabled, young and elderly individuals) [**?**].

While most acknowledge that autonomous vehicles are set to bring major reform to our increasingly overwhelmed transportation infrastructures, there remains a great deal of concern and pessimism over the legal, ethical, security and safety considerations that today remain largely without resolution.

This paper provides a review of literature into the work being undertaken to address factors found to exhibit significant influence over public trust - and ultimate acceptance - of driverless vehicles, with particular focus given within the context of the software industry as a principal contributor toward the enablement of intelligent vehicles. In this investigation, four key factors are considered: the testing and verification of autonomous systems; the redefinition of driver roles and associated impacts on passenger ergonomics; ethical considerations arising from increased reliance on autonomous decision-making; and the distinction between driver vs. manufacturer liability in the case of accidents involving autonomous vehicles.

In defining what constitutes an 'autonomous vehicle' within the context of this investigation, readers are referred to levels four and five of the On-Road Motor Vehicle Automated Driving Systems classification taxonomy published by the Society of Automotive Engineers (SAE) [**?**].

The outline for the remainder of this paper is as follows. Section 2 discusses the issue of incompatibility between existing legal frameworks relating to driver liability, and the undisputed rise in the potential for traffic accidents involving vehicles operating exclusively under autonomous control.

Section 3 provides enhanced focus on the considerations relating to the comprehensive testing and verification of autonomous vehicles, presenting work from a variety of academic and industry sources that sees software acting both as the subject and provider for new forms of testing strategies.

Considerations relating to the transfer of driver responsibility from human to computer are discussed in Section 4.

Section 5 reviews the main ethical issues that continue to present a fundamental dilemma for software engineers, as they attempt to model the complex moral reasoning required when undertaking control of a potentially hazardous craft within real-world environments.

Finally, Section 6 presents a selection of the author's own opinions pertaining to the works and ideas discussed throughout, in addition to highlighting areas where future work is likely to be needed in order to overcome remaining challenges.

II. Liability Law & Autonomous Vehicles

Whilst autonomous vehicles offer immense potential for reducing traffic-related injuries and fatalities, they will never be completely immune to the risk of crashing [**?**]. Technology faults, security breaches, and misguided behaviour in response to abnormal environments; all represent potential factors for causing an automated driver to lose control of a moving vehicle. Regardless of the form a driver may take - human or computer - the consequences of vehicle accidents remain equally damaging; costed not just in terms of financial quantities, but often - for the most severe incidents - in the loss of human lives.

Under existing legislation, liability arising from a vehicle crash is typically apportioned between two key parties: (a) the human driver(s); or (b) the vehicle manufacturer(s) [**?**]. The extent to which each party is found responsible will depend entirely on the unique circumstances surrounding each incident. In certain situations, responsibility may be attributed exclusively to the behaviour exhibited by one or more of the drivers involved (e.g. lack of adequate attention, driving under the influence of alcohol and/or banned substances etc.). For other incidents, vehicle defects may be found to represent the root cause, placing the onus on the vehicle manufacturer to compensate for loss or damage. In special cases, the evidence may find that neither party holds an enforceable proportion of the blame in light of extenuating circumstances (e.g. poor weather conditions, animals in the carriageway, vision impairment due to dazzling sunlight etc.) [**?**].

As the realisation of self-driving vehicles becomes ever more apparent, there exists growing concern amongst the legal, technical and insurance sectors as to the inadequacy of existing liability law in assessing where fault should lie in the case of an accident involving vehicles no longer under the influence of human control [**?**] [**?**]. Given that current laws surrounding vehicle liability tend to be based predominately around the notion that fault is likely to be at least in-part down to driver error [**?**], [**?**], removing this factor from the liability equation naturally raises the question of who should instead take responsibility in their place. The conclusion reached by many, sees the vehicle manufacturer become the liable party, on account of their claim of overall responsibility for the system in control of the vehicle at the time in which it crashed (most likely found to be the result of: (a) some kind of vehicle defect; or (b) a case in which the control software encountered a situation that it was not able to handle in an appropriate fashion [**?**]) [**?**] [**?**].

Whilst from a legal perspective this deduction may seem clear-cut, there is great concern over the potential for this to elicit a detrimental "*chilling effect*" [**?**] onto the future introduction of autonomous vehicles, owing to the major disincentive that it provides to manufacturers who naturally will become extremely cautious to the risk of substantially increasing the number and severity of liability cases brought against them [**?**]. For the most part, there is strong agreement that autonomous vehicle technology is set to deliver major societal benefits, and as such, work to protect and encourage innovation in this area should be viewed as an important priority for courts and policymakers [**?**]. Whilst this advocation of driverless technology is found to be shared widely across the literature studied for this review, ideas and opinions begin to diverge significantly when discussion turns to how future liability cases should be assessed so to ensure that a fair balance is struck between shielding vehicle manufacturers from an overwhelming level of personal liability, and continuing to guarantee rightful justice for accident victims [**?**].

In analysing the proposals suggested across the range of surveyed literature, two predominant themes begin to emerge: **(a)** the transference of responsibility from manufacturers to vehicle owners; and **(b)** the introduction of legislative protection for manufacturers against common liability claims involving autonomous vehicles.

A. Apportioning Liability onto Vehicle Owners

For legal cases involving an accusation of liability against another person or company, one the most critical facts that a court must determine is the extent to which the prosecution (normally the person(s) injured or otherwise affected) knew of, and voluntarily accepted, the risks involved in undertaking the activity that they wish to claim loss or damage against [**?**]. In the case where a court finds that sufficient knowledge of the potential risks was indeed held and accepted by the aggrieved party, they may subsequently elect to throw out the case through the defence of *assumed risk* [**?**].

In their paper comprehensively examining the liability implications arising from the introduction of driverless technologies, Marchant and Lindor propose that this same defence may help to provide liability protection for manufacturers of automated vehicles, by effectively reinstating the vehicle owner as the significant liable party; by account of their understanding and agreement to take on the risks associated with owning and using the vehicle on public roads [**?**]. For such a defence to become tenable, manufacturers would be required to enforce some kind of disclaimer giving evidence to fact that a buyer had willingly accepted to take ownership of the risks disclosed by the manufacturer upon agreement to purchase the vehicle. However, even with such evidence, courts are not under any obligation to accept this defence, with rulings often continuing to find in favour of the prosecution in cases where assumption of risk is raised as only a single form of defence on behalf the defendant [**?**].

Another aspect of tort law that may become critical to assessing the liability of actions made by autonomous vehicles,

is the application of rules that impose *strict liability* upon those persons who hold legal ownership over a vehicle recognised to be chattel[1] [**?**]. Under the standard imparted by strict liability, a person is held legally responsible for any damage, loss or injury that may be incurred regardless of whether or not they are directly culpable for the causes leading to such consequences [**?**]. Crucially, strict liability holds no requirement for the prosecution to prove fault on the part of the defendant; but must only convince the court that damage and/or injury was afflicted, and that the defendant was responsible at the time the incident occurred [**?**].

A popular example for illustrating the concept of strict liability focusses itself around a tiger rehabilitation centre. In the hypothetical event that a tiger escapes from its cage and goes on to inflict damage or injury, the owner of the centre is automatically held liable regardless of how strong the cage meant to contain the animal was designed to be [**?**]. A second example describes a situation where a contractor employs a sub-contractor who fails to hold adequate insurance. In the event that the subcontractor performs an action that causes either damage or injury to property or other persons, it is the main contractor who is held strictly liable for dealing with the consequences [**?**].

In examining the suitability of enforcing strict liability onto owners of autonomous vehicles, Duffy and Hopkins [**?**] look to the existing laws surrounding the ownership of an unlikely equivalent; man's best friend, the faithful dog.

Within a court of law, a dog is not viewed as a legal entity in their own right, and instead are recognised to be the personal property (chattel) of their owner, typically defined as the person in charge of and responsible for the care and wellbeing of the animal in the capacity of a domestic pet [**?**].

The implications of this are such that, in the event that a dog was found to cause harm upon another animal or person, it would be the owner who is held liable for any damages, regardless of whether or not they themselves are found to be directly culpable to the actions of their pet [**?**].

The importance of imposing strict liability in such cases, arises from the fact that dogs possess the capability to act independently of the intentions and wishes of their owners (although of course instances do exist in which the actions of a domestic animal are found to be directly attributable to unfavourable human influence (i.e. in training a dog to show aggression toward other canines or persons)). Despite this, it remains the responsibility of the owner to take appropriate measures in order to prevent such behaviour from taking place [**?**].

The defining proposal outlined by Duffy and Hopkins [**?**] suggests: as both independently-capable yet legally-unrecognised items of personal property, dogs and autonomous vehicles should be treated under the same strict liability ownership laws, as such as scheme encourages owners to take every possible precaution to the avoidance of damage or injury, given an understanding of their obligations to liability in the event than an incident occurs [**?**].

This, the authors argue, represents the fairest means of determining liability in the absence of a human driver, whilst continuing to ensure that victims are adequately protected against bearing the cost of any damage and/or injuries they may sustain.

It becomes easy for one to foresee the potential disincentive that strict liability offers to consumers, given the risk it presents for a considerable rise in insurance prices owing to the increased onus placed upon individual owners to compensate for damage costs [**?**].

However, with autonomous vehicles largely anticipated to increase traffic safety and reduce collision rates [**?**], [**?**], [**?**], many expect this will in turn work to negate potential insurance hikes on account of a re-fortified vote in confidence for driverless technology on the part of insurers [**?**], [**?**].

Although cases involving strict liability often allow for rulings to be reached more quickly - indeed, this is commonly viewed to be a strong advantage for courts handling high-volume civil litigations such as product liability or personal negligence [**?**] - owners of autonomous vehicles may still be acquitted if it can be proven that damage or injury was inflicted as a consequence of extenuating circumstances, that lie outside of an appropriate-degree of personal control; including, but not limited to, victim negligence or a serious component defect and/or malfunction [**?**].

In the papers reviewed thus far, we have seen strong advocation for protecting manufacturers from the risk of increased liability, under efforts to safeguard the future development of driverless technologies and the potential they bring to enhance vehicle safety, efficiency and accessibility. [**?**], [**?**], [**?**].

In the case of Gurney [**?**], the author raises an albeit more 'symmetric' point-of-view; proposing that liability should not simply be assigned indiscriminately to one or other party, but instead should be apportioned according to *"...the nature of the driver, and the ability of that person to prevent the accident"* whilst the car is operating under autonomous mode.

To outline their argument, the author presents four alternative case studies: *"the Distracted Driver"* - a person otherwise engaged in another task drawing their attention away from the road ahead; *"the Diminished Capabilities Driver"* - a person who exhibits a reduced ability to assume control over an autonomous driver (e.g. an elderly, young or intoxicated individual); *"the Disabled Driver"* - a person who possesses some from of disability that prevents them from driving under manual control; and finally *"the Attentive Driver"* - as a person who possesses the foresight and capabilities to prevent an accident from occurring [**?**].

By examining the unique circumstances of each case, and the relative ability of each driver to intervene in the event of a crash, the author shows that a fair assessment of liability can only be reached when these capabilities are considered under a court's ruling. For example, a disabled driver is likely to have a reduced chance of correcting the course of a careering vehicle in comparison to an able-bodied person. Therefore, in

[1] A 'chattel' is an item of movable property that is legally recognised as a personal possession of its owner [**?**].

this instance, the author advocates for the manufacturer to be held liable, given that the vehicle has been deemed suitable for use by persons who would otherwise be unable to operate a manually-driven vehicle due to their disability. In a more complex case involving a distracted driver, the author proposes that it would be reasonable for a court to expect such a person to intervene in preventing a collision, given sufficient and adequate warning by the system as to the impending failure of the autonomous driver to maintain control [**?**].

Whilst this approach allows for a tailored assessment of liability to be drawn for each independent case, there are some potential problems that must be considered also. First and foremost, in the case of a liability claim involving an autonomous vehicle, under what evidence could it be proven that an occupant exhibited behaviour attributed to that of an attentive driver, rather than the case a defence is likely to put forward in that the driver was distracted? Although in-car video recording may contribute as evidence in such cases, this raises significant questions over the issue of occupant privacy [**?**], and the extent to which this evidence could be found to give sufficient corroboration as to the driver's ability to assume control when alerted to warnings given by the vehicle's software.

Secondly, by requiring courts to examine the specifics of each claim on a case-by-case basis, lawsuits involving autonomous vehicles can naturally be expected to place increasing pressure on the civil court system as the number of cases rises with the rate of adoption for the technology [**?**]. It is for this reason, that both Marchant and Lindor [**?**] and Duffy and Hopkins [**?**] suggest imparting liability on the vehicle owner in the default case prevents courts from having to invest their resources into determining the exact cause in each and every case brought before them.

In all of the above cases, insurance for autonomous vehicles has been viewed predominantly as a disincentive for owners - a part of the problem, rather than the solution. Schellekens [**?**] takes an altogether more favourable view, proposing to distribute the cost of liability damages across the wider population of autonomous vehicle owners through a system of *"obligatory insurance [...] where the duty to insure rests on the holder of the vehicle."*. To give support and context to their argument, they refer to existing legislation adopted in Sweden, which enforces a rule of obligatory first-party insurance to cover strictly for damages incurred by accident victims [**?**]. Under such practices, compensation for the victim is given precedence over determining liability [**?**] - the driver's insurance is always expected to pay out for damages to the victim, before attempting to claim back any costs through civil proceedings at a later date [**?**]. While this idea to place victims at the forefront of consideration is a noble one, enforcing a policy of obligatory insurance is unlikely to bring much appeasement to consumers already displaying scepticism toward the likelihood of increased costs, as previously highlighted in [**?**], [**?**] and [**?**]. However, the author holds the firm belief that *"...this effect - if it occurs at all - [would be] counterbalanced by diminished legal expenses."* [**?**], but of course, only time will tell how significant this argument will become in light of the extent to which driverless cars make true of their promise to decrease collision rates.

B. Legislative Protection

For law makers tasked with determining responsibility in the case of driverless vehicle accidents, a significant dilemma stems from the contention that both parties exhibit toward accepting liability, and the implications that this raises toward the potential development and adoption of autonomous vehicles. On the one hand, manufacturers show a distinct level of trepidation and cautiousness toward placing themselves in a position of increased liability by fronting development in the technologies necessary to make driverless vehicles a reality. On the other hand, consumers are unlikely to show interest or willing to invest in a technology which either: (a) elicits them personally responsible for damages caused from their autonomous vehicle going awry; or (b) raises insurance costs to a point where autonomous vehicles no longer represent value for money.

An alternative approach that may help to reduce these liability concerns, would be the introduction of legislative protection for manufacturers, to either limit, or remove the degree of liability that could be brought against them [**?**]. This would in effect continue to leave manufacturers as the party held primarily responsible in the majority of cases, but prevents prosecutors from collecting excessive payouts on account of the manufacturer's culpability [**?**]. Whilst such an approach would likely appease some of the apprehensions raised on behalf the automotive industry, new issues begin to emerge over the negative impacts this may have on the degree to which manufacturers will remain incentivised to push for improvements in product safety under efforts to reduce their own liability [**?**]. Therefore, it will become essential that such measures are taken with a view to find an optimal balance between stimulating continued innovation and maximising public safety.

Although legislative intervention is not taken lightly for protecting individual technologies and industries [**?**], statutes such as the Oil Pollution Act (1990) [**?**] and the Price-Anderson Nuclear Industries Indemnities Act (1957) [**?**] represent examples of where federal legislation has successfully been introduced to defend organisations and industries from suffering an intolerable degree of financial burden as a consequence of major disasters or incidents. Prominent examples include the $71 million awarded to help cover the costs of claims following the 1979 *Three Mile Island* partial nuclear meltdown [**?**]; and the cap of $75 million placed upon damages payable by BP following the *Deepwater Horizon* explosion and subsequent oil spill back in 2010 [**?**].

Whilst the benefits of legislative protection are evident in instances where the potential exists for widespread industrial bankruptcy, awarding dedicated protection for manufacturers of autonomous vehicles is liable to set a precedent for comparative technologies that raises some difficult and contentious questions regarding where law makers should 'draw the line'

for protection coverage. For example, if laws were to be passed giving reduced or withdrawn indemnity to manufacturers of autonomous vehicles, would it be reasonable to expect the same protections to be applied to cases involving non-autonomous vehicles? Should fault found to be on account of a vehicle defect be viewed as any less severe when software is control of the wheel vs a human driver? Answers to such questions will only become clear as time and discussions continue to move forward.

III. Verifying Safety of Autonomous Vehicle Software

Driverless vehicles present unique and difficult challenges for software development and testing teams on account of their increased reliance upon inductive inference and high-level reasoning for allowing them to operate safely within uncontrolled, unstructured and often unforgiving dynamic environments [**?**].

With autonomous vehicles set to be placed in an unparalleled position of trust over the safety of their occupants, other road users and public bystanders, the need for rigorous and uncompromising verification of software integrity and safety becomes unequivocally essential. However, the sharp rise in complexity and behavioural obscurity exhibited within these systems leads many to believe traditional software-safety techniques will be rendered largely inadequate to the requirements stipulated in order to present driverless control systems with any chance of attaining safety certification [**?**], [**?**], [**?**], [**?**]. To prevent this from becoming a potential barrier for future development of autonomous systems, work to find new approaches to verifying the integrity and reliability of autonomous software is becoming an area of increased research focus.

Wagner and Koopman present two contrasting yet equally pertinent approaches for realising more effective methods of assuring software safety in autonomous vehicles. In their first paper, the authors advocate Karl Popper‘s notion of *falsificationism* as a means of treating the act of testing more as a scientific endeavour rather than simply a “find-and-fix” development exercise [**?**]. It was Popper‘s belief, that a scientific theory only becomes meaningful if the potential remains for it to be later disproved given supporting empirical evidence [**?**]. This idea acts extend the long-held understanding that whilst to conclusively prove a theory one must show evidence for it holding in *all* feasible cases, conclusively disproving this theory requires the demonstration of only a single negative case [**?**].

The authors describe how this notion of *“denying the consequent”* has the potential to flip the traditional view of testing completely on its head. No longer is testing regarded as an exercise focussed on gathering *“mountains of confirmatory evidence”* for findings of correct behaviour in nominal cases, but instead looks to actively seek out instances evidencing a contradiction of expected behaviour, and in so doing, directing developer attention toward addressing these - often overlooked - behavioural edge cases [**?**].

To translate this high-level philosophy into employable testing practice, the authors propose the use of run-time verification in order assess the extent to which an autonomous system upholds the expected outcomes of its associated safety case [**?**]. First, a modified version of metric temporal logic - a popular extension of linear temporal logic for providing specification and verification of real-time systems - is used in conjunction with state-machine descriptions to encode requirement claims and safety-case expectations into a temporal model of system-state rules [**?**]. The system is then examined at run-time to monitor for any unanticipated behaviour not recognised when compared against this formal specification [**?**]. This, the authors note, not only serves as a means of detecting *“subtle cracks”* in the integrity of software safety cases, but also can contribute to mitigating against the risk of safety violations arising as an unintended consequence of software systems that exhibit high code and behaviour complexity [**?**].

The application of falsificationism to software testing appears to show promising potential, effectively liberating the tester from a role focussed on continuously proving the already expected and/or known, to instead becoming the driving force behind the proactive discovery and rectification of worst-case behaviours. Of course, applying these approaches is likely to require significant effort and careful thought on the part of both developers and testers, however, as Wagner and Koopman are keen to highlight - *“...there is tremendous benefit in even just pondering how to formally specify safety properties of a complex system”* [**?**], by helping to promote safety maximisation as a priority for developers.

In a later paper, the same authors take their analysis one step further by exploring the challenges associated with the testing and validation of autonomous vehicles within the context of existing automotive safety practice. They introduce the ISO-26262 international standard, an automotive-specific variant of the IEC-61508 *Functional Safety of Electrical/Electronic/Programmable Electronic Safety-related Systems* standard [**?**], which provides vehicle manufacturers with a standardised process used for developing life-critical automotive control and Engine Control Unit (ECU) software [**?**].

At the centre of the ISO-26262 standard, lies the ‘V-model’ software development framework (see Figure **??**) [**?**]. An extension of the ‘Waterfall’ development framework [**?**], the ‘V-model’ describes a sequential development and validation process in which each development phase (represented on the left-hand side of the model) holds a direct association to a corresponding validation phase (represented on the right-hand side of the model). The process begins with the requirements analysis, before moving down through the design and into implementation representing the bridge between the two sides. As the process rises up the right-hand side, the focus of testing widens from validation of individual classes and components, up to a broader verification of the system as a whole through a combination of integration and acceptance testing [**?**].

Whilst this process provides manufacturers with a stable footing for adopting robust software safety and verification

practices, the authors discuss how the progression into fully-autonomous driving systems presents a number of challenges in *"mapping the technical aspects of the [autonomous] vehicle to the V approach"*. In total, the authors identify five key technical aspects that at present remain inadequately covered by the existing approaches recommended as part of ISO-26262: *"driver out of the loop"*; *"complex requirements"*; *"non-deterministic algorithms"*; *"inductive learning algorithms"* and *"fail-operational systems"* [**?**]. In each case, the authors find their cause to be attributed to one or a combination of: an inevitable rise in system complexity; and the demand for more stringent assurances on system stability, security integrity and fault accountability [**?**].

Potential solutions that can be applied generally to help mitigate these types of issue include: **(a)** adopting a strategy of incremental deployment for autonomous driving operations in order to limit the scope of requirements under consideration at any one time; **(b)** employing use of a 'monitor-actuator' paired architecture to provide segregation of complex autonomy operations from the more basic - but none-the-less indispensable - safety assurance logic; and finally, **(c)** introducing fault injection techniques across the range of system hardware and software tiers (such as Fuzz Testing [**?**] or Compile-Time Injection [**?**]) in order to help expose the more obscure edge-cases that - by their definition - are rarely triggered as part of a standard testing strategy [**?**].

In their conclusion, Koopman and Wagner highlight that whilst many challenges remain on the path to reaching a comprehensive mode of safety-certification for high-level autonomy, it would not be beyond the bounds of possibility for systems deployed for autonomous vehicle control to be architected in such a way that would allow them to become assessed under the kinds of existing safety practices employed in industry today [**?**].

This notion appears to make a large degree of sense, given that standards like that of ISO-26262 and others represent the culmination of significant time, effort and empirical evidence to arrive at a set of recommended practices found to provide acceptable safety assurances. It would therefore seem injudicious to throw all of this pre-existing knowledge to one side, especially when many - if not all - of these recommendations can continue to be applied to software developed for autonomous vehicles following the comparatively small amount of work required to address the kinds of challenges and incompatibilities highlighted in [**?**].

Fisher et al. take a contrasting view toward the verification of general autonomous systems than what has been offered thus far. It is their belief, that in order to undertake the most effective evaluation of high-level autonomy behaviour, it is important to focus on assessing the *intentions* of the system drawn upon its own beliefs about itself and the environment it occupies, rather than on the real-world outcomes that may or indeed *may not* be directly attributable to the actions a system has chosen to take [**?**].

Critical to this assessment, is the ability to capture these high-level reasoning concepts within some type of formalised representation structure. In the work of Fisher et al., this role is fulfilled by the prominent *Belief-Desire-Intention* (BDI) model for rational-agent architectures [**?**], in which the authors define a 'belief' to represent *"the agent's (probably incomplete, possibly incorrect) information about itself, other agents and it's environment"*; a 'desire' to represent *"[one of] the agent's long term goals"*; and finally an 'intention' to be *"a goal that an agent is actively pursuing"* [**?**].

Having obtained a BDI-based 'specification' stipulating the required behavioural properties for an autonomous system, the authors then combine this with standard mathematical and logic-based methods to perform an adapted kind of formal verification[2] to assess the extent to which the system matches the expected behaviour profile whilst operating within *"an unrestricted environment representing the real world."* [**?**].

Whilst the authors spend much of their time delving into the technical details of their verification approach (later publishing a follow-up paper providing a fully-comprehensive explanation of the methodology and its intricacies [**?**]), it would seem that the real value of this work comes from the philosophy that it fosters. In any life situation, there always exists the possibility, that the *intended* outcome of an action or sequence of actions will fail to correspond with the actual turn of events. This observation is no less applicable in the case of autonomous beings as it is in humans; for example, whilst an autonomous car may *intend* to avoid hitting pedestrians at all conceivable costs, there always remains the potential situation that a child may run out into the road before the car has time to take evasive action. If this situation was to fall as part of an evaluation assessment, would it be right for the system to be found unfit for purpose, given that its *intention* was to always avoid pedestrians and take evasive action where required to try and ensure this?

The overarching argument made by Fisher et al., is that this would not represent a justifiable approach to verifying the high-level reasoning behaviour implanted within an autonomous system [**?**]. It seems reasonable to assume that it will never be possible to fully model each and every potential situation and outcome that may arise from having an autonomous system operating within real world environments. Therefore, rather than obtaining a limited and often inaccurate assessment of the behaviour of a high-level autonomy drawn upon the effect it has within the real world, testing and verification methods should instead be looking to ensure - to the greatest possible extent - that the *intentions* of these systems meet our desired expectations of safe and considerate reason, and that the beliefs used to derive these intentions give as accurate a view of the real world as is possible [**?**].

This idea begins to break down the boundary between the *functional* and *ethical* verification of an autonomous system, when we consider that the ultimate objective of a high-level autonomy should be to make decisions that achieve a specific task, whilst - crucially - continuing to uphold the same moral

[2]Formal verification describes the process of employing mathematical and logical methods to evaluate the behaviour of algorithm against the expectations stipulated as part of a pre-established specification [**?**].

values accepted and followed by humankind. Whilst it seems insurmountable for testing and verification to ever eradicate the possibility of an autonomous system undertaking an unsafe action, the work conducted by Fisher et al. gives credence to the notion that it will be possible for us to assure ourselves that the systems we entrust with our safety will never *deliberately* choose an action that is *more likely* to cause us harm or inconvenience than is presented by any of the other choices it has available at the time [**?**].

IV. Vehicle System Security

Although the rise in in-car autonomy is set to offer a plethora of significant and potentially life-changing benefits to both existing road users and those who currently remain unable to drive, increasing attention given by cyber-criminals toward the exploitation of in-car security vulnerabilities represents a powerful deterrent to the public acceptance of driverless technologies [**?**], [**?**], [**?**].

Whilst today's vehicles remain capable of demonstrating only a very limited range of semi-autonomous functionality (i.e. automatic braking, adaptive cruise control, automatic-parking etc.), the findings from growing research in the exploitation of in-car entertainment, cellular and diagnostic systems make for worrying reading.

In a joint-published review of automotive cyber-security risks [**?**], the Institution of Engineering and Technology (IET) and the Knowledge Transfer Network (KTN) highlight a number of academic studies focussing specifically on highlighting vulnerabilities in the security of electronic automotive systems dating back to 2010. However, this study has found evidence of consideration toward the cyber-security of vehicles arising much earlier than than 2010, with Koopman writing in 2004 about the security risks posed through providing vehicles with an internet connection [**?**]; a feature that has since gone on to become a major USP for vehicles in today's market [**?**], [**?**].

In the IET/KTM paper, their review of academic research projects begins by introducing the work of researchers from California-San Diego and Washington Universities, who successfully intercepted and later disrupted internal Electronic Control Unit (ECU) communications through the use of custom software exploiting access gained through a cable connection to the vehicle's diagnostic port [**?**]. This represents one of the first projects shown to successfully override critical locomotive and safety components (including the engine, door locks and window controls) by transmitting messages over the Controller Area Network (CAN) bus used to enable the numerous ECUs present within the car to communicate with one another [**?**] [**?**].

In the following year, the same team of researchers presented results from experiments, obtained both on and off road, that demonstrate their ability to remotely override and track a modern mid-range vehicle through an extensive variety of interfaces including compact discs, radio channels, cellular and Bluetooth connections [**?**]. Examples of implemented exploits include installing a *"CD-based firmware update [to] re-flash the media player ECU"* [**?**], allowing for custom overrides to be initiated when a particular station name is received over an FM signal; using the in-car cellular connection to download a small C-based IRC client exploit for passing CAN bus messages sent via any internet connection to the car to execute; and employing an *"Android phone and trojan app"* [**?**] to gain access to the internal CAN bus via a buffer overflow attack [**?**].

In 2014, Valasek and Miller published a comprehensive paper detailing their success in using approaches similar to those highlighted in [**?**] and [**?**] to override critical vehicle components in the case of two different vehicles produced by independent manufacturers [**?**]. Their work acquired significant media attention, after they demonstrated the ability to assume control of functions including the steering, information displays and even the brake pedal, all whilst the vehicle was in motion and under the direction of a human driver [**?**] [**?**]. In addition to confirming the findings of previous work [**?**], [**?**] along with introducing a range of new attacks, the authors use their work to highlight the need for greater accessibility for researchers in conducting these types of investigation in a way that supports open and transparent discussion between academia and the automotive industry [**?**]. To this effect, the authors announced plans to release all of the tools and data gathered in their investigation into the public sphere for others to use - an act never before undertaken as far as current evidence suggests [**?**]

Since this time (and beyond the scope of the IET/KTM survey [**?**]), Valasek and Miller have published another research paper extending on their previous work, in which the authors show not only the ability to assume near-total control of a vehicle featuring the very latest electronic systems and security measures, but to do so entirely from a remote location, and without requiring any kind of alteration to the original test vehicle [**?**].

In this case, the authors were able to obtain access through a vulnerability exposed within the internet-enabled in-car entertainment system known as the *"head unit"* [**?**]. This system was found to have direct access to both of the primary CAN buses used to inter-connect the vehicle's numerous ECUs. By sending messages over a cellular connection to an insecure IRC port monitored by the head unit, the authors found themselves able to remotely transmit CAN packets to the vehicle, and in turn, infiltrate the normal operation of the various safety-critical and non safety-critical systems [**?**].

Exhibiting their findings through a provocative public demonstration involving subjecting a journalist to a remote attack whilst driving along a US highway [**?**], the author's work received soaring media coverage [**?**], [**?**], [**?**], and prompted the vehicle manufacturer in question to initiate a recall over 1.4 million vehicles in response to their findings [**?**].

All of these findings present a very bleak outlook for the security, and ultimate safety, of fully autonomous vehicles, that are intrinsically set to depend on electronic systems to an entirely different degree to that of even today's most sophisticated vehicle makes and models.

Whilst greater openness and transparency from vehicle

manufacturers is an area identified by many in the security industry as requiring drastic and immediate improvement [**?**], [**?**], [**?**], [**?**], what is striking about all of these investigations, is the evidence they provide for indicating a lack of appropriate authentication applied by vehicle manufacturers to protect electronic systems (including safety-critical systems) from unauthorised communications. On a similar point, Valasek and Miller propose that it is possible for vehicle systems to 'detect' many of the harmful communications sent by hackers through inspection of traffic patterns recorded over a CAN bus network [**?**]. They find that under normal operating circumstances, CAN packets sent between ECUs tend to follow a marked and repeatable timing pattern, causing their own custom packets to become exposed owing to the irregularities exhibited in their transmission behaviour [**?**]. The authors also note how a sudden and unexpected proliferation of diagnostic CAM packets (found to only traditionally exist in the case of authorised system inspections) may also contribute to highlighting the presence of a unauthorised intruder packets that are often characterised as diagnostic messages [**?**].

Other mitigation strategies proposed for future vehicle systems include employing redundant control systems to help avoid the risk of a single point of failure across safety-critical components [**?**], [**?**]; supporting a "voting-logic" approach derived from the "Byzantine general" problem, to coordinate the defence against the possibility of an attacker gaining control over all redundant systems via a single point of access [**?**]; and finally enforcing the need to consider and formally assess cyber-security requirements as part of existing requirements analysis [**?**], [**?**].

V. Summary & Conclusion

This paper has provided an introductory - but by no means comprehensive - study into a selection of critical considerations that are set to hold significant influence over the viability of future fully-autonomous driving technologies. From the literature studied under this review, it has become clear that the potential advancements this technology can offer for vehicle safety, accessibility and efficiency are to be signficant. However, many concerns remain over the verification, security and liability implications that driverless vehicles present, as they usher in a new wave of cyber-physical transportation possibilities.

In the legal field, a battle is brewing between manufacturers and consumers as to who should be made financially responsible in the event that a vehicle under autonomous control becomes embroiled in a traffic collision. Legislators face a difficult task in balancing the concerns of both parties, so to avoid the risk of deterring investment in a technology that promises so many far-reaching social benefits. Whilst wholesale legislative protection provides incentives for industry to continue development and innovation of driverless technologies, it comes at the price of increased taxes for individuals and the potential for relaxed attitudes toward iterative improvements on system safety and reliability to emerge, on account of the reduced liability risk for manufacturers. In striving for an optimal balance between stimulating continued innovation and maximising public safety, the adoption of specialist 'driverless' insurance, in combination with potential exemptions for disabled or otherwise incapable drivers, would appear to provide a strong foundation from which further discussions can begin.

Further evidence in support of this approach is noted in the announcement made by the UK Government (May 2016) for plans to create a new 'Modern Transport Bill', set to *"...put the UK at the forefront of autonomous and driverless vehicles ownership and use..."* by *"encouraging potential investors in autonomous vehicles"* and *"ensuring appropriate insurance is available to support the use of autonomous vehicles."* [**?**].

As vehicles continue to depend ever more upon increased autonomy and high-level decision making, developers responsible for creating these autonomous systems will need to consider carefully how to sufficiently verify their behaviour in order to assure the authorities, and the wider public, that they can be trusted with ensuring our safety. The complex and often obscure behavioural nature of autonomous systems renders many traditional testing and verification practices inadequate. This requires the software industry to find new perspectives toward the assessment of autonomous behaviour, diverting focus away from the outcome of individual actions, to instead understand and verify the intentions and beliefs that drive a system to take certain choices.

It is at this point we realise, how the values we hold over our own ethical conduct are likely to become a rising influence over the judgment we pass for verifying the behaviour of high-level autonomy. It therefore becomes our obligation, to ensure through further research and collaborative efforts, that we can adequately define, represent and verify the upstanding of 'ethical principles' within the context of an independent yet safety-critical autonomous control system.

Of course, the integrity of driverless systems remains only through their capability to defend against corruption. As security researchers have shown, hackers and cyber-criminals prose a real and worrying threat not only to the security of a vehicle's electronic systems, but ultimately to the safety of its occupants and others in the vicinity.

Whilst it is clear that vehicle manufacturers have some way to go in order to sufficiently demonstrate that their systems can deliver a robust defence against malicious attacks, existing research in this area shows promise that public confidence can be recovered, if industry is prepared to exhibit greater cooperation and transparency with researchers to implement appropriate initiatives to enhance the security of future vehicle systems.

Unfortunately, this review has been unable to cover all of the many factors set to influence consumer consideration toward the acceptance of autonomous vehicles. Concerns over vehicle cost, passenger ergonomics, driver 'handover' practices and 'moral' decision-making continue to sit alongside the legal, testing and security considerations that both the industry and society in general must overcome, if a 'driverless' future is to ever become a true reality.

In order to achieve this, we must endeavour to treat the qualitative analysis of these considerations with the same vigour and dedication as ongoing technical innovation, through undertaking systematic reviews of continued research evidence and conducting a comprehensive meta-ethnography [**?**] of consumer opinions to extrapolate further insights.