Analyzing Wariness of Autonomous Vehicles Through the Vehicle Decision-Making Process

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Quinlan Dawkins

Spring 2020

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Sean M. Ferguson, Department of Engineering and Society

Introduction

The rise of autonomous vehicles (AVs) has become a major discussion in both technological and political spheres. As the technology has developed, safety concerns became more prevalent to match the severity of the consequences posed by a crash. And because an algorithm is handling a crash scenario, defining the correct behavior in these scenarios is a part of these safety concerns. Part of the problem with defining safety concerns stems from trying to understand the standards surrounding AVs. Here, standards refers not just to safety requirements but also to the general set of requirements and expectations associated with an autonomous vehicle. Varying definitions of automation, varying implementations, and a range of governing approaches have left some potential future consumers, investors, and regulators wary of joining the space (Marshall, 2020).

Standards are a powerful tool for understanding new technologies and in turn, limiting the variation that has caused wariness. Challenges facing AVs can be broken down into technological, ethical, environmental, and legal challenges where the most controversial topic relates to the technological problem of designing the vehicle decision-making process. Furthermore, standards analysis of a decision-making system can be categorized by the sensing system, client system, action system, and human-machine interface (Martínez-Díaz & Soriguera, 2018). Due to the nature of standards analysis, in this paper, the categories of challenges, ethical, environmental, and legal, are first briefly explored. This context is used to explore the technical aspects of a vehicle decision-making process from a standards perspective, both in terms of the standards it faces and the ones created by it. This includes breaking down the analysis based on each component of the decision-making system.

Standards as a Framework

Before analyzing literature, the standards analysis framework needs to be defined. The framework for standards analysis is provided in "Standards and Their Stories" (Star & Lampland, 2009). Standards are described as being an attempt to simplify what is often a 'messy' reality. As a result, they are often necessary. Beyond necessity, they tend to have a set of commonalities as well. First, they are nested inside of other standards. The standardization of placing an order on Amazon is linked to the standard of owning a credit or debit card which is linked to the standard of having a bank account. Second, they can be distributed unevenly. The ability to prepay a speeding ticket means a rich person might bear less risk when speeding than a poorer person. Third, their impact depends on context. Making hiring decisions based on a college degree will by design impact some differently than others. They tend to be integrated into the sociotechnical systems we face daily. The standard of having a phone number or an email address is depended on by each of the online accounts we have (e.g. Facebook). Finally, standards can be a manifestation of the ethics and values of the system they are standardizing (Star & Lampland, 2009).

With these properties in mind, using standards as an analysis framework for research of a sociotechnical system requires placing the standards inside of some sort of context. This starts with defining the infrastructure under which the sociotechnical system and its standards are developed. Infrastructure is defined as the parts of a sociotechnical system that is invisible to the user. A key point is that the infrastructure for one system might itself be a sociotechnical system (e.g. highways). This suggests that infrastructure might be nested similar to standards.

Additionally, depending on the system, the infrastructure can be hard to define, and once defined, determining how it affects the use and standards will still depend on context. Blind and deaf people face street infrastructure in a sometimes completely different way from the average person on a sidewalk. Thus, to analyze a sociotechnical system based on its standards, we need to place it in some sort of relevant context, define how actors interact with the system to define the infrastructure and then in turn, define the standards for the system. From here, the standards are applied to a system to see how they influence decisions in and around the system.

Literature Review

The Society of Automotive Engineers (SAE), a standards setting organization that defines technical and engineering standards, first defined five levels of automation in 2014 ("J3016: Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems", 2014). Furthermore, the National Highway Traffic Safety Administration (NHTSA) has adopted this definition as it has become a standard way to understand levels of automation ("Preparing for the Future of Transportation: Automated Vehicle 3.0", 2018). These definitions pertain to what the system is able to do without driver input with no reference to actual implementation.

Challenges Facing Automation

Many of the challenges facing automation in vehicles, as defined by the literature reviewed here, can be broken down into the categories of technical, legal, environmental, and ethical. In this section, the latter three categories are explored for use in the analysis of a decision-making system. Not every challenge cleanly fits into one category, nor are these categories comprehensive. Rather, they are adequate for capturing the essence of where AVs deviate from traditional systems, as well as providing relevant context for the analysis of the technical part. Additionally, in this context, challenge does not refer only to problems AVs must overcome to be viable, but also opportunities for improvement over current systems. As an example, consider the claim that the most controversial challenge facing this technology today pertains to the vehicle decision-making process (Martínez-Díaz & Soriguera, 2018). This relates to three of the categories of challenges, where there is a technical challenge of implementing a decision-making process, the ethical challenge of defining the process, and the legal challenge of standardizing and enforcing a justified approach.

Regarding legal challenges, in 2014, a revision was made to Article 8 of the 1968 Convention on Road Traffic, an international treaty aimed to improve traffic safety. The change removed a requirement for a driver to be active at all times and allowed for autonomy in the driving process (Barabás, Todoruţ, Cordoş, & Molea, 2017). This change shows legal bodies' awareness of AVs, but is not the whole picture. Ambiguity in the true decision-making process for an AV poses a challenge for defining legal requirements, as does the legal issue of liability in the event of a crash (Martínez-Díaz & Soriguera, 2018). Additionally, challenges such as how strict to be relative to human drivers, licensing, and data privacy all need to be given definitions under law (Zhao, Liang, & Chen, 2017).

In an attempt to understand the current state of safety standards for AVs, I expand upon the legal attempts that have been made to reconcile automation with existing traffic infrastructure. A report from RAND from 2018 suggests that there is no industry-wide standard definition of safety with regard to AVs (Fraade-Blanar, Blumenthal, Anderson, & Kalra, 2018).

This does not mean work has not been done in this space, however. That same report, as well as another from University of Michigan, make independent attempts to define frameworks for measuring safety of an autonomous vehicle (Peng, & McCarthy, 2019). Furthermore, the U.S. Department of Transportation published a report in late 2018 that describes how communities can prepare for the future of transportation, particularly concerning AVs ("Preparing for the Future of Transportation: Automated Vehicle 3.0", 2018). Finally, we have seen states begin implementing legislation pertaining to AVs in 2017 and 2018, with twenty-nine states and Washington D.C. passing some sort of legislation ("Autonomous Vehicles | Self-Driving Vehicles Enacted Legislation", 2019).

Looking at the current state of safety standards, as of 2019, in California, 52 companies hold testing permits for self-driving cars on roads (Hancock, Nourbakhsh, & Stewart, 2019). Unfortunately, it's hard to gather significant statistical information on the testing that happens on these vehicles. AV field tests still underway have not had statistics published and DMVs do not publish separate crash statistics themselves, making it difficult to build a large enough data set for statistical analysis in an area where total volume of testing is dwarfed by daily human hours driven (Wang, & Li, 2019). A report from this year gives statistics on 113 crashes to try and understand the current safety of these systems (Wang, & Li, 2019). Another problem is that safety testing has minimal standardization outside of the permission and crash reporting process due to the wide variation in degree of automation across autonomous vehicles (Hancock et al., 2019). This is part of why a black box approach is often chosen when evaluating AVs from both a legal and ethical standpoint.

Environmental challenges reflect the sentiment that if there is an opportunity to be efficient for the environment, it might as well be made a requirement. When considering the problem of congestion on roads, there is environmental motivation for solving the problem, where additional congestion on roads could lead to increased emissions (Bagloee et al., 2016). Other environmental opportunities include limiting the intensity of fossil fuels due to more control of the vehicle's usage on the manufacturing end, lighter vehicles due to optimizing safety through avoidance over damage control, and general fuel efficiency, as an algorithm should be much better at maximizing miles per gallon than a human (Bagloee et al., 2016).

Finally, consider the ethical challenges. Closely tied to the legal challenge of providing a legal definition for the requirements of a decision-making algorithm, when considering the ethics of an AV, scholars often look to the "trolley problem" as a way to find a moral dilemma in the design of an AV (Martínez-Díaz & Soriguera, 2018). Attempts to resolve ethical, as well as legal, challenges almost invariably are approached by treating the car and its algorithms as a black box (Martínez-Díaz & Soriguera, 2018).

The Infrastructure of Trust for Self Driving Vehicles

To conclude the literature review, we begin to unpack part of the infrastructure these standards operate in. At a most basic level, this includes physical infrastructure defined by the likes of roads and traffic signs, but the more relevant is the infrastructure of trust surrounding transportation. When navigating some sort of sociotechnical system, trust is what allows a user to simplify many complex decisions, some of which rely on missing data. In the case of driving, one example of trust is that the vehicle will stop when the brakes are applied, or accelerate when the gas pedal is pressed (Tong, 2016). Furthermore, few complex systems can be used without some level of trust. Now, considering road traffic, there is trust built into the road safety laws and regulations that allow drivers to avoid worrying about whether a nearby driver will stop at a red light.

With this idea of trust in mind, we can look to other literature to understand how the trust infrastructure built into existing traffic systems has translated to self-driving technology. Tong argues that what defines a safe car for a specific person depends on the degree of trust they have in the black box vehicle and driving safety networks. Furthermore, this trust is in part defined by being able to control the result, where someone is able to gain trust through predictability and control over the system (Tong, 2016). Through this framework, the intuition is then that self-driving technology begins to break down what was able to build up this infrastructure of trust in the first place. A study on user acceptance and experience with self-driving cars suggested with survey data that automation has a negative correlation with acceptance and experience. This correlation is reversed if a driver had experience with Advanced Driver Assistance Systems (ADAS) however (Rödel, Stadler, Meschtscherjakov, & Tscheligi, 2014). Additionally, a more recent study on Austrian consumers surveyed consumers on their feelings towards various aspects of AVs. They were able to show the opposite; positive correlation between a feeling of safety and the level of automation. They showed a negative correlation between concerns about AVs and a willingness to use the technology as well (Wintersberger, Azmat, & Kummer, 2019).

Vehicle Decision-Making

Focusing now on the problem of vehicle decision-making, Martínez-Díaz and Soriguera define four main technical elements that most self-driving car designs comprise: the sensing system, client system, action system, and human-machine interface, where each of these parts pertain to a set of technical challenges. The list of technical challenges includes environment perception, short term path planning, longer term path planning/map integration, and vehicle control mechanisms (e.g. speed, direction) (Zhao et al., 2017). Each of these challenges relate to the elements of an autonomous system, and each one has multiple solutions. For example, radar perception vs. laser perception are both possible solutions to the problem of environment perception (Zhao et al., 2017). In addition to the challenges pertaining to proper driving, there is a challenge related to communications between autonomous vehicles. The previous technical challenges were internal challenges to allow integration with human drivers, but if every vehicle on the road were significantly less aggressive than a human driver, as early prototypes have been, an estimated 600 vehicles per hour per lane would be lost (Martínez-Díaz & Soriguera, 2018). This congestion problem is estimated to be an opportunity for improvement, however, through reduced congestion from accidents, enhancing vehicle throughput, and reducing total miles traveled per trip (Bagloee, Tavana, Asadi, and Oliver, 2016). This challenge manifests as a technical challenge of inter-car communication.

In the following sections, each of the components of a vehicle decision-making system is analyzed. It is important to note that the functionality of each component varies based on the level of automation in the vehicle. As a result, automation levels as defined by the SAE are part of the standards for each of the following components. Additionally, we will see that the challenges discussed in the literature review are a standardization of the functionality of each part. Furthermore, applying the infrastructure of trust to each of the components allows us to break down how wariness has manifested in the space of autonomous vehicles.

Sensing Systems

A sensing system is responsible for collecting the necessary environment data. Among the most frequently used sensing technologies are LiDAR or laser perception systems, radar, and camera systems (Zhao et al., 2017). Combinations of these sensors are often used based on the strengths of each technology in terms of accuracy and range. Fundamentally, however, all of these sensor types are attempting to solve the same problem: collecting data on the surrounding environment for the other parts of the decision-making system to output vehicle actions and possibly information for the driver, depending on the level of automation. The accuracy required by the complete sensing system increases directly with the level of automation, but the standard for sensing systems is well-defined by this problem.

With this in mind, how do we map this standard onto wariness. When considering human drivers, the "sensing system" equivalent are the human sensory systems, primarily vision and hearing. Legally, drivers licenses represent that a specific individual's senses are acceptable for safe driving. This, however poses a challenge for regulators, as designing safety thresholds for drivers licenses depends on an expectation that the experience with visual and auditory information is similar between people. Furthermore, people are willing to trust their own senses, hence an added layer of trust is required due to the automation of sensory systems. This gives a natural interpretation of how wariness is affected by sensory systems, and for both regulators and drivers that should be a question of accuracy and reliability. This goes beyond simple vision distance testing done at the DMV and is closer to testing highly varied environments in an industrial setting.

Client Systems

The client system is required to take and process the information from the sensing system (perception task) and transform it into an action for the vehicle to perform (decision task). This component of the decision-making process is the most complex, the least standardized, and the most controversial. Both tasks are naturally computation intensive, especially if employing deep learning techniques. As a result, there is some variation in the hardware approaches considered by some manufacturers. Full compute boxes with processors and accelerators, system-on-chip solutions, and cloud computing based solutions all either exist in some capacity or are possible extensions of existing hardware (Martínez-Díaz & Soriguera, 2018). In reality, client system hardware can be effectively black-boxed as all solutions can be modeled with an internet connection and a representation of compute power. This is useful as representing compute power, both in terms of a quantitative measure and in terms of the tasks that can be performed (e.g. video playback), has already been standardized in other computer systems.

Moving on to the perception task, there are two main methods that have been developed for visual perception: Simultaneous Localization and Mapping (SLAM) and machine learning based image recognition and environment reconstruction (Zhao et al., 2017). In short, SLAM reconstructs the environment by comparing incremental changes in sensor data over time. The details are out of the scope of this paper, but the important point is that both methods are designed to produce a representation of the environment for use in the decision task. If we take this to be a standard, we can effectively isolate the decision task as the major cause of wariness in AVs. This is in part due to the fact that there is a measurable level of accuracy in this, as well as the other components discussed. From a consumer, investor, or regulator’s perspective, measurable accuracy translates nicely into existing engineering processes and allows for testing to be well-defined. Safety ratings are based on minimizing risk by black-boxing a system and defining an acceptable range of outputs. The standards described here show that outside of the decision task, no component or solution prevents that approach.

This leaves the decision task. Martínez-Díaz & Soriguera describe it as one of the most challenging tasks an AV must perform. It encompasses prediction, path planning, and obstacle avoidance. A key point is that this is where the correct action in "awkward" situations needs to be defined. This ties into the ethics problems highlighted in the literature review and shows why the decision task likely is the greatest cause for wariness with regard to AVs.

Action Systems and the Human-Machine Interface

The action system takes an action from the client system and physically performs it. The action system is to a client system what a steering wheel and pedals is to a human driver. Because this system has such a direct analog in a non-autonomous vehicle, the standards for this system are much better defined than for a client system. As a result, the action system doesn't contribute to the wariness of regulators, drivers, or investors any more than similar systems do in conventional vehicles.

The human-machine interface (HMI), however, is a critical part of the experience with an AV for the consumer. In the same way drivers are willing to trust that pressing the brake will slow the car down, as well as putting trust in the accuracy of the speedometer, the interface in an AV is closely tied to the trust of the driver. The degree of input from the driver decreases with the level of automation, but the kind of information displayed to the driver is not necessarily well-defined, especially when considering level 5 automation. What we see is that because SAE level-4 vehicles and beyond are generally unavailable (Martínez-Díaz & Soriguera, 2018), current autonomous vehicles retain the necessary equipment and displays for normal vehicle operation. Thus, it can be reasoned that the familiarity of standard HMI elements serves to break down the wariness faced by consumers.

Conclusion

By many accounts, self-driving cars are the future of transportation. How we transition to this future while feeling safe is another question entirely. What can be seen through this research is that while people might be wary of AVs as they stand today, the standards for the most controversial part of these vehicles, the decision-making process, is generally not far removed from non-autonomous vehicles. Sensing systems, action systems, and the HMI all either have a direct analog for a human driver or represent a machine version of what a human normally does. For the client system, the perception task still fits into existing safety standardization methods, isolating the decision task as the largest cause of wariness in AVs. It should be noted that there are a number of other technical factors in the development of AVs, such as path planning, that wasn't covered in depth in this research but likely does play into the hesitation consumers, regulators, and investors exhibit when it comes to autonomous vehicles.

References

Barabás, I., Todoruț, A., Cordoş, N., & Molea, A. (2017). Current challenges in autonomous driving. *IOP Conference Series: Materials Science and Engineering*. https://doi.org/10.1088/1757-899X/252/1/012096

Basl, J., & Behrends, J. (n.d.). *Why Everyone Has It Wrong about the Ethics of Autonomous Vehicles*. NAE Website. Retrieved February 16, 2020, from https://nae.edu/221212/Why-Everyone-Has-It-Wrong-about-the-Ethics-of-Autonomous-Vehicles

National Highway Traffic Safety Administration., U.S. Department of Transportation., (2017). Automated Driving Systems: Voluntary Safety Self-Assessments; Public Workshop. Retrieved from Federal Register website: https://www.federalregister.gov/documents/2017/10/17/2017-22058/automated-driving-systems-voluntary-safety-self-assessments-public-workshop

National Conference on State Legislatures., Autonomous Vehicles | Self-Driving Vehicles Enacted Legislation. (n.d.). Retrieved from http://www.ncsl.org/research/transportation/autonomous-vehicles-self-driving-vehicles-enacted-legislation.aspx

Fleetwood, J. (2017). Public Health, Ethics, and Autonomous Vehicles. *American Journal of*

Public Health, *107*(4), 532–537. https://doi.org/10.2105/AJPH.2016.303628

Fraade-Blanar, L., Blumenthal, M. S., Anderson, J. M., & Kalra, N. (2018). Measuring Automated Vehicle Safety. Retrieved from https://www.rand.org/pubs/research_reports/RR2662.html

Hancock, P. A., Nourbakhsh, I., & Stewart, J. (2019). On the future of transportation in an era of automated and autonomous vehicles. *Proceedings of the National Academy of Sciences of the United States of America*, 116(16), 7684–7691. https://doi.org/10.1073/pnas.1805770115

SAE International (2014), *J3016: Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems*. Retrieved March 6, 2020, from https://www.sae.org/standards/content/j3016_201401/

SAE International (n.d.), *SAE Standards Development: Standards Development Process*. Retrieved March 6, 2020, from https://www.sae.org/standardsdev/devprocess.htm

Logue, K. D. (n.d.). *Should Automakers Be Responsible for Accidents?* 6. Retrieved from https://www.cato.org/sites/cato.org/files/serials/files/regulation/2019/3/regulation-v42n1-1.pdf

Marshall, A. (2020). Investors Hit the Brakes on Automotive Startups. *Wired*. Retrieved March 5, 2020, from https://www.wired.com/story/investors-hit-brakes-automotive-startups/

Martínez-Díaz, M., & Soriguera, F. (2018). Autonomous vehicles: Theoretical and practical challenges. *Transportation Research Procedia*, *33*, 275–282. https://doi.org/10.1016/j.trpro.2018.10.103

Peng, H., & McCarthy, R. L., (2019). A Concept to Assess the Safety Performance of Highly Automated Vehicles. Retrieved from https://mcity.umich.edu/wp-content/uploads/2019/01/mcity-whitepaper-ABC-test.pdf

Preparing for the Future of Transportation: Automated Vehicle 3.0. (2018, September 27). Retrieved from US Department of Transportation website: https://www.transportation.gov/av/3

Rödel, C., Stadler, S., Meschtscherjakov, A., & Tscheligi, M. (2014). Towards Autonomous Cars: The Effect of Autonomy Levels on Acceptance and User Experience. *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 1–8. https://doi.org/10.1145/2667317.2667330

Star, Susan Leigh & Martha Lampland (2009) ‘Reckoning with Standards’, in M. Lampland and S.L. Star (eds), Standards and their Stories (Ithaca: Cornell University Press): 3–24.

Tong, J. W.-S. (2016). Do you trust your car? Enacting car safety. Retrieved from https://pdfs.semanticscholar.org/7e6e/d500bc48fc4d9300713f8007edcc7e85386a.pdf

Wang, S., & Li, Z. (2019). Exploring the mechanism of crashes with automated vehicles using statistical modeling approaches. *PLoS ONE*, 14(3). https://doi.org/10.1371/journal.pone.0214550

Zhao, J., Liang, B., & Chen, Q. (2017). The key technology toward the self-driving car. *International Journal of Intelligent Unmanned Systems*, *6*(1), 2–20. https://doi.org/10.1108/IJIUS-08-2017-0008