2017

Assigning Liability in an Autonomous World

Agni Sharma
Claremont McKenna College

Claremont Mckenna College

Assigning Liability in an Autonomous World

Submitted to
Professor Adrienne Martin

By
Agni Sharma

For
Senior Thesis
Fall 2016
12.05.2016

Contents

I. Introduction

In the event that an accident occurs on the road, the standard response (once ensuring the safety and well-being of the people involved) is to exchange insurance details and confer with witnesses to determine a causal link that faults one, if not all drivers involved. This system of assigning liability has been the historical standard, and is actively followed today; but what about in the future? What if the cars involved in the accident at hand are driverless cars – autonomous vehicles (AVs)? An account of the accident can still be given, but it is now more difficult to determine fault – to whom do we follow the causal link. Can we still apply our established liability principles in such an autonomous world, and if so to which party? This study explores the question of liability in an autonomous world, discussing potential frameworks of implementation to AV accidents, and attempting to fill any gaps with our currently used liability system. The paper will start by presenting the trolley problem as a sample of decision-making scenarios that AVs may face; and will highlight the significant boost in road safety that is a result of AV implementation. Section II aims to build an understanding of vehicle autonomy, discussing the need for machine ethics and its requisite ethical conditions, as well as the standard for autonomy in vehicles ; followed by the main discussion on possible liability frameworks to be adopted.

Trolley Problem

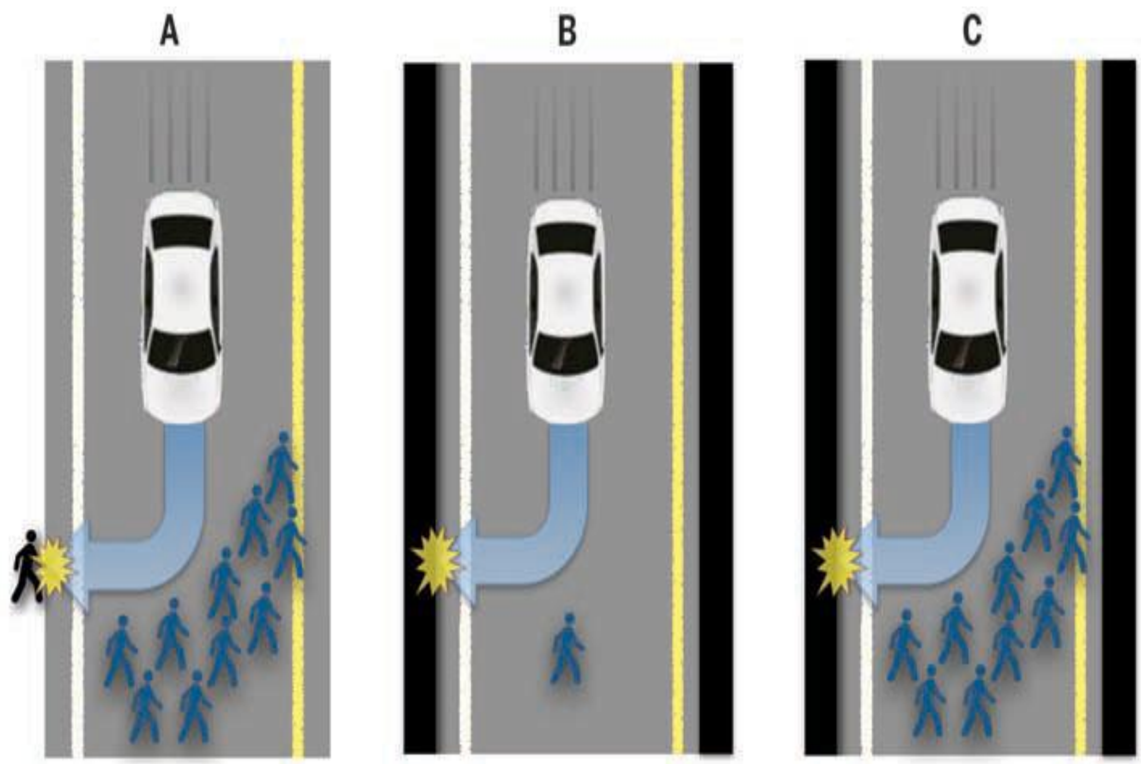
One of the most famous philosophical problems is the ethical dilemma faced in the trolley problem. The trolley problem has been long discussed and debated and has been slightly adapted in a number of ways, but remains unchanged in its fundamental ethics. Introduced by Philippa Foot, the problem is as follows:

A trolley rounds a bend and is now progressing toward a portion of track that is being maintained by five railway workers; who do not have any space around the track to move to. The driver is unable to stop the trolley, which will surely collide with and kill the five workers. You, a bystander, happen to be standing by a lever which will reroute the trolley to a track on the right; however, that track has one railway worker on it, who would be killed. Should you pull the lever to save the five, but kill the one? The answer seems to vary among individuals, but there is a general consensus that no moral violation occurs should you decide to pull the lever. Consideration of this ethical dilemma is extended by a second scenario, usually introduced as a follow-up placing you on a bridge with opportunity to throw a 'fat man' onto the tracks to derail the train and save the five workers at the expense of the fat man. There seems to be greater consensus that pushing the man would be morally impermissible, despite the higher number of deaths that occur as a result. Using the two scenarios the trolley problem indicates that individuals' moral intuition and judgement varies even in scenarios that produce outcomes that risk, as well as preserve the same number of lives.

Although the consideration of AVs does not invoke the ‘problem’ that is differing moral intuitions to a similar outcome, AVs will encounter scenarios such as the first half of the trolley problem. The question of which path to take, and subsequently how much damage to cause, must be answered in the moment by the AV. The AV’s capability to provide an answer is in the hands of engineers and programmers participating in the development and manufacture of the vehicle. The AV must make a decision on a course of action – that is to kill five or one – to take based on the principled outcomes that are programmed into it.

The programmers are removed from the direct personal responsibility that comes with physically pulling the lever, but must face the decision of whether the lever is to be pulled or not; since there must be some precaution in the event of inevitable damage, even though AVs are poised to reduce road deaths. The process of producing a feasible AV requires manufacturers to undertake morally significant decisions. This decision making must be translated and programmed into the vehicle; and includes consideration of the number of lives to save, as well as the sensitive decision of whether to prioritize passenger safety more highly than the safety of non-passengers.

A study titled *The social dilemma of autonomous vehicles*, conducted by Jean-Francis Bonnefon, Azim Shariff, and Iyad Rahwan gathered information through six online surveys. The social dilemma presented is similar to the trolley problem, except that if the car saved five pedestrians, it would swerve and kill the passenger/driver. The first survey asked participants to assess the morality of programming AVs to risk the driver and passenger, and of the 182 participants, 76% believed it was moral to save ten pedestrians for the sacrifice of one passenger. The second survey used a larger sample size of 451 participants, asking for moral approval for a varying number of pedestrians between one and a hundred. There seemed to be the belief that a single pedestrian should be sacrificed, but moral approval for the driver-sacrifice programming increased with the number of pedestrian lives put in danger; and grew to become consistent with the 76% approval rate acquired in the first survey. Looking at image 1 below it is clear that participants would prefer option C and even option A, to the unfavorable option B.



[1]

[1] Ackerman, *People Want Driverless Cars with Utilitarian Ethics, Unless They're a Passenger.* See references

Participants seemed to prefer a utilitarian framework for the decision making system inbuilt in the AV[2], but deviated from this preference when presented with the decision to preserve themselves or a loved one as a passenger – at the cost of non-passengers.

This variation in decision making amongst potential consumers of AVs suggests that manufacturers will have great difficulty determining which moral intuitions are programmed into the vehicle. The lack of clarity in determining an ideal moral framework that enables vehicle decision making may provoke the question, why continue producing AVs if we can't identify the perfect set of morals? Although assessed from a largely utilitarian standpoint, AV crash statistics suggest significant reductions in annual road deaths, a decrease of about 30,000 deaths; and such significant improvement is likely to be wholly supported by all road commuters as well as society as a whole. It is rational to want further production and development of AVs given this information. The following section illustrates the current trends of road accidents and the prospective impact of the introduction of AVs.

Crash Statistics

Prior to further discussing liability concerns, it will be useful to take a look at crash statistics to better understand the road accident climate in the United States. The statistics presented first will not include any testing AVs currently active on roads. The most recent reports available are for the year 2014 and are presented below:

[2] Keep this in mind during the section on AI and machine ethics, as the importance of a utilitarian foundation is established, suggesting that utilitarian concerns should not be discarded.

In 2014 a total of 29,989 fatal crashes were reported, taking 32,675 lives; while an additional 2,300,000 people were injured in vehicle crashes. Table 13 below provides further insight into the distribution of these accidents.

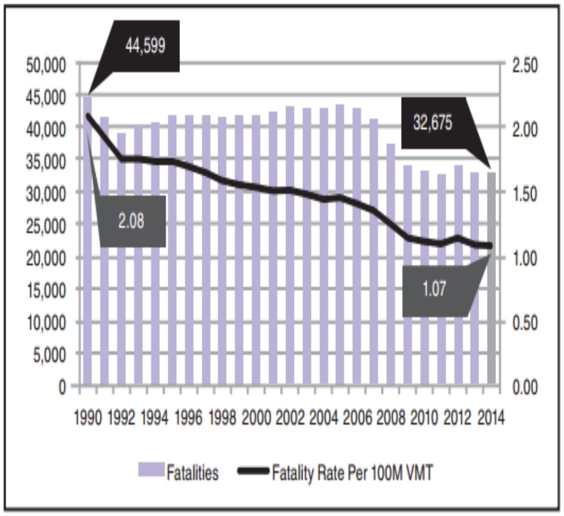
	Motorist		**Non-motorist**
Drivers	16,454	**Pedestrians**	4,884
Passengers	5,751	**Pedal Cyclists**	726
Unknown	71	**Other/Unknown**	203

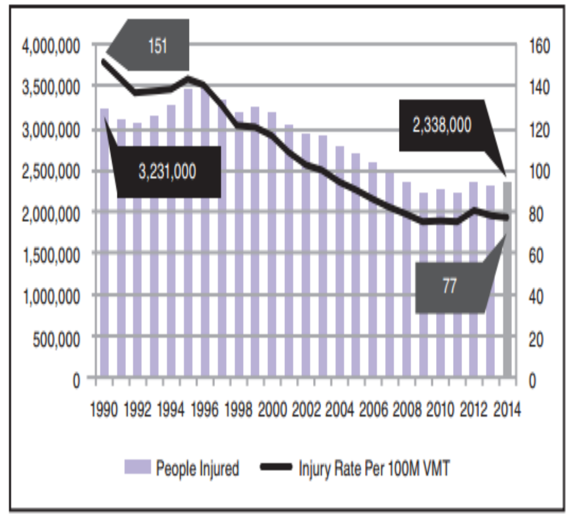
The US Department of Transportation grouped fatalities by presumed cause and presented these findings:4

- 31% (9,967) of total fatalities were caused by alcohol-impaired driving
- 10% (3,179) of total fatalities were caused by distracted driving
- 2.6% (846) of total fatalities were caused by drowsy drivers
- 28% (9,262) of total fatalities were caused by speeding-related crashes.

The image below shows historical trends in crash fatalities, fatality rates, injuries, and the injury rate.

[3] "National Statistics" FARS-NHTSA. See references.
[4] "Traffic Safety Facts – Crash Stats." *US Dept. of Transportation – NHTSA*. See References.





Source: FARS 1990-2012 Final File, 2013-2014 ARF.

This information accentuates an important piece of knowledge – approximately 72% of road accident fatalities occur primarily due to some form of human error, which largely defines the current format of liability processes. In fact Chris Urmson, the director of Google's self-driving car program, reports that up to 94% of crashes are caused by driver error5. Urmson also reports that in about six years of testing Google's AVs on the road, the test vehicles have been involved in eleven minor accidents that caused nothing more than light vehicular damage; out of which the AV is said to have caused none of the crashes. The drastic change suggests the immense use and utility that can be gained from AVs, minimizing human error. Further, out of eleven accidents, eight incidents were rear-ending crashes – scenarios that will be reduced by highly sensitive technology built into the AVs.

The information presented in this section makes a strong case for pushing past decision making and programming hurdles to proceed with the production of AVs. The

[5] Urmson, Chris. "The View from the Front Seat of the Google Self-Driving Car."

margin by which AVs are expected to reduce road accidents and deaths emphasizes the value of the vehicles, hence encouraging a rapid transition from human driven vehicles to AVs, within the bounds of safety.

II. Framework for Autonomy in Vehicles

The notion of autonomy is the most critical distinction between the current and future state of vehicles; and it is the implications of this notion that create the need to restructure the liability framework. Thus, it is necessary to analyze and understand autonomy – what does it mean for a vehicle to be autonomous? The term ‘smart car’ has increasingly found its way into our vocabulary, as it becomes a common prefix for more capable technology – smartphone is the most common example, while the development of other household machinery ushers in new terms such as smart fridge, smart tv etc. Machine capability continues its rapid climb, with more intelligent devices and appliances arriving annually. However, these machines are still not considered autonomous. Perhaps then the presence of Artificial Intelligence (AI) is needed for autonomy – after all, it does allow machines to act independently, with minimal user input. Common examples are bots such as Siri and Alexa. It is useful here to develop a better understanding of what exactly AI is and how it serves a purpose in our discussion of liability.

The most important aspect of AVs is the notion of *autonomy* – which, put in simple terms, is enabling the machine at hand to become an independent decision-making entity. In order to make decisions the machine must be programmed with some degree of intelligence. Such a process of imparting knowledge on machinery falls in the realm of artificial intelligence. The general worldwide view of artificial intelligence is skewed by long-standing fantasy and hyperbole in the entertainment world, as well as speculation of the danger of machine sentience. Hence, the definition this paper will consider is somewhat broad: artificial intelligence is simply the notion of development of computer systems to mimic certain human abilities, such as speech recognition, relevance perception, analysis of data in the immediate environment, and learning for better decision-making. Once again considering the example of Siri, although the bot appears highly-intelligent, and bears some of the mentioned traits for AI functioning, it is still not considered autonomous. Here, questions of the threshold between intelligence and autonomy are provoked. These characteristics of vehicles must be defined to better understand the capabilities of AVs.

The term "intelligence" tends to be broad, referring to varying levels of development. The word itself is a semantic placeholder for any entity with comparatively higher capabilities. To streamline the discussion, we resort to the following definition, "Intelligence is the ability to formulate one or more action sequences which can increase the probability of successfully achieving the system's goals in an uncertain environment." (Gunderson, Gunderson 141). Let's apply this definition to a set of vehicles which we'll call intelligent vehicles (IVs). The system here is the IV's AI computer, and its goal is to evade accidents, or to cause the minimum damage in an inevitable accident. An IV will survey the situation at hand and study it to develop a list of courses of action, but leaves it to the driver to select the preferred course. Although the IV conducts cognitive activity, there is no resulting path of execution determined by the technology within the IV itself. This unrestricted ability for execution would allow an IV to be considered an AV.

In order to execute a course of action, the AV must be able to compare possible courses and determine the best outcome. Determination of the outcome is based on the core values programmed into the AI system. Similar to the human process of selecting courses of action based on a set of inherent rules and morals, the vehicle must be given an ethical guideline by which to make to decisions. The need for a guideline introduces a key concept – machine ethics – which grapples with the ethical dilemmas of programming such human traits into computer systems. For a greater understanding of the importance of machine ethics in any conversation on artificial intelligence, Tonkens provides some clarity:

> The nascent field of Machine Ethics is gaining momentum. Much of its fuel stems from the perceived imminent and inevitable (Allen et al. 2006, p. 13; See also Sparrow 2007, p. 64) development of artificial moral agents (hereafter AMAs), who will be able to (or already *do*) perform morally consequential actions in the world. Because autonomous machines will perform ethically relevant actions, akin to humans, prudence dictates that we design them to act morally.
>
> (Tonkens 2)

Although Tonkens terminology of artificial moral agents is not followed throughout this study, it provokes discussion of the standard required for any ideas and concerns to be considered as part of the machine ethics concept. Discussions of machine ethics are intended to assess the acceptability of autonomous machinery in real-world situations; or in other terms , "*machine ethics* is concerned with the behavior of machines towards human users and other machines." (Anderson, Anderson 1). AVs simply taking action that results in positive ethical outcomes is not considered to come under machine ethics, which focuses on the inherent code of ethics upon which the vehicle exercises its autonomy. Looking at an example: robot jockeys can replace young enslaved boys in camel races in Qatar, freeing them from slavery6. Although this may seem an ethical act by granting the young jockeys freedom, it does not require the programming of an ethical framework into the machinery, and hence is not considered to fall under machine ethics. With this notion of a code of ethics we can analyze autonomy and how it's exercised.

Humans and a number of other members of the animal kingdom are granted an autonomous status due to the attribute of self-governance. Since the discussion of liability adds a tinge of morality to ideals of autonomy, we can further explore self-governance through a Kantian lens. Kant argues that this notion of moral self-governance requires an adherence to one's own self-imposed rules, rather than an obedience to any set of externally imposed rules. It remains true, however, that individuals may willfully abide by external principles without detriment to their autonomy. Although tethered to this moral anchor, humans individually interact with these morals under their own self-imposed set of rules.

While autonomy in humans seems to be guided by a code of ethics and laws, it does not bear the exact same meaning as autonomy in the case of AVs. The degree of autonomy programmed into the vehicle must achieve desirable outcomes, such as meeting the standard of social acceptability – maximizing benefits while limiting risks.

---

[6] Example taken from Anderson, Anderson; provided by James Moor, symposium keynote speaker.

The vehicle's autonomy is rooted in the program and code of ethics written by the manufacturer, which provokes the question of what code of ethics do we use? Surely ethical principles are not collectively consistent within the population of a township, much less so with the entire nation's citizen body.

The ethics programmed into the vehicle is reliant on the code of ethics that manufacturers believe to be apt, which is likely to be an extension of a set of morals that add to the market value of AVs. Given the subjectivity in individuals' ethical virtues, it is possible that ethical flaws can be found in any outcome that directly results from the vehicle's ethics. Programming the vehicle to save passengers, to sacrifice passengers, or to cause any property or bodily damage involves a moral decision that could be challenged by different ethical perspectives. Due to the severity of societal consequences that come with poor AV decision making, it is imperative that the code of ethics of the vehicle meets some acceptable social standard. We can refer to this standard as a social good or a social optimum. Reaching optimality entails accessing the full benefits of AVs, which in turn brings minimal risk. If a specific ethical framework more effectively prevents accidents, the overall risk of using the road is reduced; thus it is likely to be socially acceptable. Socially acceptability is the standard that must be strived for, as it is an indicator of optimality. It is likely that this set of principles will be declared as ubiquitous moral maxims to be programmed into AVs.

Given a liberty to self-govern comparable to that of humans, vehicles can ascend into autonomy. An IV and AV share systemic goals and both rapidly produce solutions, but AVs progress a step further by sifting through solutions to find the optimal option, and executing the best course of action. “Autonomy is the capability of systems to select between multiple possible action sequences to achieve the system’s goals, based on the current situation and internally defined criteria.” (Gunderson Gunderson 141). The term ‘internally defined criteria used in this definition alludes to the functioning of the AI system; specifically focusing on the moral and decision-making framework programmed into the vehicle.

The ultimate course of action chosen by the vehicle’s AI is based on the set of rules given to the system. Decision making is a necessity for vehicle autonomy, and it is important for manufacturers to consider a decision making framework that attempts to achieve the system’s goals so that greater societal goals can be met. Achieving these goals requires AVs to be in constant coordination, that is – the vehicles must agree on intended courses of action and be able to communicate and analyze information on the next planned movement to ensure safe transition. The necessary coordination of AVs suggests that the moral framework – which guides decision making – programmed into the vehicle has to be one that contains some universal aspect that enables AVs to work in unison.

This objective framework is the only external set of rules imposed. The vehicle maintains its ability to self-govern; the only alteration being that its decision making process is rooted in an external framework. The vehicle does not have to go outside its own AI system to analyze and execute courses of action.

In order for this autonomy to be enabled the vehicle must have the ability to accept input data from the environment around it. Humans exercise autonomy based on sensory input, and similarly manufacturers must build such capabilities into AVs. Due to the magnitude of the risks that come with vehicle use, sensors on AVs should provide in-depth information, beyond just motion and distance sensing. The AV sensors must collectively provide enough data to determine the number of obstacles on the path (as well as those on alternate paths), and differentiate between types of obstacles, among other capabilities. The AV needs maximal information to select and execute the best possible outcome.

With this established understanding of vehicle autonomy and its requirements we can move forward to discussing the benefits of AV autonomy. The most significant advantage that comes with autonomous capability in vehicles is the elimination of complete driver control. The state of the driver is uncertain and unpredictable; as erroneous behavior can be induced in people by a plethora of factors. The introduction of autonomous capability allows the vehicle to evade the risks tethered to human error, and in turn significantly reduce the overall risks of commuting by road.

Another benefit that comes with vehicle autonomy is the additional utility afforded to users by foregoing the need for active user control of the vehicle. Without the need to dedicate complete focus to road and traffic conditions, the AV user can make better use of the time spent commuting. Users avoid the fatigue of driving, furthering the potential for productivity. Additionally, autonomous vehicles unlock a new consumer base, providing individuals with an inability to drive with access to self-sufficient transport.

Equipping a vehicle with an autonomous system will however bring an inherent set of risks. AV users have to trust the system in all decision-making situations, which involves acceptance of the risk of the vehicle making a wrong decision. It seems rational to be more averse to enabling autonomy if it brings more risk. The potential consequences of making a wrong decision are so severe that it raises questions of whether risking the uncertainty of autonomous decision making is worthwhile. To justify taking such risk, the impact of AVs on the level of road safety must be reemphasized. The capabilities of AVs are set to reduce road accident and death by a substantial margin, which suggests that enabling the autonomous capabilities of vehicles reduces the overall risk of road travel. This is to say that although AVs have some small probability of making the wrong decision, the vehicles have to run the risk of making that decision on fewer occasions. The superior level of road safety brought about by enabling vehicle autonomy limits and outweighs the risks that come with decision making errors, justifying granting vehicles the relevant autonomous capabilities.

To produce the greatest potential benefit from AVs, manufacturers have the important task of molding the vehicle's autonomy to more effectively avoid risks. This threshold of autonomy should allow the vehicle to maintain decision-making capabilities, but limit the potential courses of action to the most rational and acceptable outcomes. The vehicle's autonomy must be constrained by the appropriate set of road, transport and vehicle laws that push for an optimal social outcome. Additionally, the vehicle's decision-making system must not produce outcomes outside of predominant moral standards. It is imperative that the vehicle's autonomy remains rooted in the framework programmed in by the manufacturer, to prevent the decision-making from being influenced by individuals' moral standards. AV users should not have the ability to alter the vehicle's autonomy to serve personal purposes, as it can impact coordination by causing incoherence in communication. The system shouldn't be able to find and exploit loopholes in the existing legal structure. Further, consideration of liability will be impacted by the capabilities of the vehicle and the foundational factors involved in its decision-making.

Ensuring the optimal decision-making by AVs facilitates the replacement of human drivers, creating an environment with less risk. However, despite the improvement in general safety, there remains some probability that AVs will encounter an accident. The potential for an accident, however small, indicates that over time a non-negligible number of accidents is likely to occur. The occurrence of accidents ushers in questions of liability. Damages caused by accidents must be accounted for, provoking further considerations of how and to whom liability is distributed. The framework for assigning liability must account for a number of considerations. The process involves a reallocation of individuals' property based on the circumstances surrounding the accident in question; thus accounting for a purely objective tally of fatalities and damages is incoherent with this ideal. Additionally, to fully achieve the benefits of AVs, we must avoid any obstacles to the production of, or demand for the vehicles. Another consideration is of justice – liability must maintain fairness and uphold acceptable moral standards.

So why not just fall back on insurance and liability processes that are ubiquitously followed today? The following section will present and analyze the currently prevalent model of liability to determine its feasibility in an autonomous world, using its principles to fill gaps in other possible frameworks.

III. Solutions

The Purpose of Liability

Thus far we have made some mention of liability and the need of a novel, re-worked system for its assignment; but before analyzing such potential frameworks an overlooked question must be addressed – what is the purpose of a system to assign liability? Understanding the purpose of liability provides an ideal towards which all participants in the AV realm must strive to achieve. It is the basis upon which we guide our thinking and frame potential liability models.

Liability systems aim to preserve a generally acceptable moral standard of road sharing, that is upheld by enforcing a level of moral responsibility on all parties. This aspect of liability allows for progression toward the optimal social good by targeting manufacturers and drivers as agents, exacting a cost on both finances and conscience to improve future decision-making and the resulting actions. Liability channels agency in a manner that is coherent with societal benefit and incentivizes progressive betterment of accident preventive behaviors or processes.

Let us first consider the implications of a liability free system. Subtracting the threat of liability eliminates accountability[7] for actions. This elimination occurs not only for drivers, but also for all parties that could be involved in an accident – including pedestrians and even the manufacturers. Without any accountability, we are left with no

[7] This accountability will be primarily financial and legal.

way of regulating behavior, which could allow drivers and manufacturers to act purely with self-concern, as there is no consequence for doing so. To put in another manner, the absence of liability leaves no incentive to further the social good if it conflicts with personal goals. Although achieving the desired level of social good will provide universal benefits – such as a lower probability of road accidents – individuals may still choose to forego these benefits for more immediate and direct advantages; and must be encouraged to prioritize social concerns.

It is valuable to justify our assumption that agents will indeed succumb and act according to self-interest. Observing everyday driving practices, we can rationally derive some degree of self-interest over social concern. Consider speeding on highways as an example. Given an uncongested highway, numerous vehicles on the road tend to travel at speeds over the legal and safe limit. Drivers speed in order to more efficiently achieve their goals, by arriving at their destination faster. In addition, the relative risk of speeding is low, as fines are of low-cost. If the liability costs of all vehicle behavior is lowered to a similar extent, it is likely that drivers would more greatly risk road safety to achieve personal goals. To better clarify this argument we need a macro-level example – hence we will consider the traffic patterns in India.

Traffic in India is notorious for its unruliness and erratic flow. Although the state of Indian traffic laws is on the rise, there has historically been improper enforcement of the laws, as well as an absence of some necessary rules – lane discipline in particular. Due to the lacking enforcement, drivers are more willing to take risks under the belief that they are likely to escape any liability; a willingness that translates to daily driving behavior. There exist scenarios in which drivers will operate their vehicles under the influence of alcohol, knowing the immense risks and safety hazards, because of the same belief of avoiding liability. The impact of lacking lane discipline laws is most prominent in traffic patterns, as rarely is lane discipline observed. Given the absence of the threat of liability, drivers shift freely and unpredictably between lanes to expedite their commute. In the event of an accident, drivers often do not resort to official means, and choose to negotiate privately.

The possibility of private negotiation is another obstacle to achieving the purpose of liability. If people do drive recklessly, without a system of liability we may not know of it. People may settle disputes privately, or just drive away from an accident since there is no burden of liability. The problem with this is that in the case of AV accidents, information will not be relayed back to manufacturers – information that is needed to improve the AV's functioning. If we do not know what is going on, we can't hold manufacturers to standards of development, as they won't know all the details of what went wrong. It is possible that AV users too will have less incentive to act according to guidelines over self-interest in a number of scenarios. Liability systems must be put in place to curtail the self-prioritization of all parties that could be involved in an accident.

Viewing liability as a form of sentencing allows us to break down and analyze liability through the lens of Hallevy[8]. The purpose of any form of sentencing could follow under the following areas: retribution, incapacitation, deterrence, and

[8] In *When Robots Kill*, Gabriel Hallevy studies and analyzes criminal sentencing of AI entities, ultimately arguing for theories of application of these sentences to corporations, specifically those behind the AI entities in question.

rehabilitation. Retribution is a notion that does not fit the motives of auto accident liability. The basis for retribution is a form of negative reinforcement, exacting some suffering on the perpetrator, to pacify society by counteracting the damage caused. Liability under retribution will serve as a retaliation to the offense committed, ignoring any effect of consequential action on the social benefit. AVs generally lack the cognitive processes that enable suffering to influence humans, thus applying such punishment to the vehicles is futile. Since retribution is achieved only through suffering and no other means, it cannot be applied to AVs; hence we take the notion no further.

Incapacitation involves physical measures to ensure that the violation is not committed again, and is present in more extreme cases of accident liability. Such inhibition for humans could include financial constraints, withdrawing of driving privileges, or even forms of physical confinement like house arrest and imprisonment. Although such methods are effective for humans, it seems impractical to attempt to inhibit AVs in the same manner. We can however go a step further and apply forms of incapacitation to the manufacturers, as AV agency is an extension of the manufacturer's. This corporate imprisonment model will be discussed later under manufacturer liability.

Deterrence and Rehabilitation seem to be the central concerns for any sentencing system – including auto accident liability. The threat of liability costs should be sufficient to discourage and prevent actions that may cause accidents in the future, while also motivate a change in the mindset of the liable party in the accident. Deterrence encapsulates the first half of the previous sentence, pushing for a means of discouraging future violations. It takes a two-pronged course of action: (a) deterring the agent committing the action (b) deterring others from committing the action. Although the liable agent faces the greatest impediment to committing future violations, others are averse to being in the position of the liable agent; and are indirectly discouraged from committing punishable activity.

Rehabilitation goes a step further in attempting to change the mindset of perpetrators. The notion of rehabilitation is applied in a number of present day scenarios, with both physical and mental concentrations. Taking alcohol as an example, the purpose of rehabilitation is to develop an inherent aversion to alcohol within the user. Similarly, the ideal liability system must foster a mindset that is averse to further violating any laws. With an established understanding of the purpose of liability we can analyze and compare different frameworks to paint a picture of what accident liability could look like.

Current Framework

Any discussion of potential frameworks must consider the presently predominant model. Our analysis of possible solutions must first ask the question of whether there exists a need for a novel solution, if a feasible and functional framework is already in place. In order to appropriately respond to this question, we must first look at the adequacy of the current system for assigning liability to determine the need for any efforts put into creating a fresh framework.

The current and long-standing framework of auto-accident liability is one that emphasizes the causal network around the accident. Stating this in a simplified manner, we need to know what events or behaviors actually caused the accident. Liability follows the causal link, ultimately being pushed onto the individual(s) that committed the relevant actions and errors. The process of following a causal chain most commonly results in an assessment of driver fault as the principal determinant of liability assignment. The justification here is that the driver of a vehicle has committed some moral flaw if his own negligent actions lead to the harm of others, or the damage of property. There is however a twist in the currently employed model – the causal chain may not necessarily link back to just a single party. Taking an example, a driver A may have been startled by negligent actions of another driver B, or by an illegal maneuver by a bicyclist or pedestrian. Under such circumstances, the entire burden of fault is not borne by driver A; and can be additionally divided between driver B and the cyclist as well. Tangible versions of this example are presented in the form of case law later in this section.

It will be useful to delve into the notion of fault under this framework, as better understanding of how fault is determined will paint a clear picture of the process of liability assignment. Common law segregates fault into three primary subsections: negligence, recklessness, and intentional misconduct[9]. Although these terms are further broken down for legal specificity, a broader understanding is considered for the purposes of this paper. Negligence is commonly seen in cases of driver fault, and is simply carelessness while operating vehicles. All operators of motor vehicles are required to exercise a certain degree of care that will minimize any predictable risks and afford maximum protection to other patrons of the road and immediately surrounding environments. Negligence can occur in both active and passive forms: an example of the

[9] "Fault and Liability for Motor Vehicle Accidents." *FindLaw*. See References.

former being something such as not yielding the right of way or running a red light, and the latter something such as not noticing a red light or another vehicle in the blind spot due to distracted driving. While negligence can be a passive act, recklessness is defined as a "willful disregard for the safety and welfare of others."[10] An action is considered to be reckless behavior if the driver has an understanding of the possible damage and harm that may occur from the action, but makes the decision to commit the action regardless of the presented concerns. Examples of reckless behavior include driving under the influence of alcohol, or other sensory impairing substances, and driving well above the known speed limit. The active understanding of the risks that come with the committed actions merits greater liability being placed on the driver than in cases of negligence.

The third subsection, intentional misconduct, goes one step further. An act is considered to be intentional misconduct if the driver not only understands the risks that come with committing an action, but pursues the commitment of that action with the intent of actualizing the potential risks and harm. This type of fault is rooted in the mindset of the driver; and an example of this could be a premeditated hit-and-run. Intuitively, the burden of liability that falls on the driver will be the greatest amongst the three subsections under fault.

With this established understanding of the current framework of liability assignment, we can begin to determine its worth in the autonomous realm. Can we simply adopt the same framework in essence? Let us start by considering the dominant theme of driver fault within the system. For better analysis, the notion of driver fault should be

[10] Ibid.

tweaked just slightly, being understood as human fault – determined by human error. Although human error will remain somewhat of a concern, its significance will greatly diminish as AVs intend to minimize such error; and hence we can assume that a higher number of accidents will occur as a result of issues outside of human shortcomings. In light of this challenge, the assignment of fault becomes complicated due to a now obstructed causal link.

It is interesting to note, however, that despite such a crippling challenge to the current framework, themes of negligence, recklessness and intentional misconduct can contribute in a similar manner to whichever new framework is implemented. To better understand this implementation under new systems, we can hypothesize applications to cases in which the driver seems not to be at fault. In such situations, fault based models can be applied to other parties surrounding and involved in the accident in a similar manner to the driver. The process includes assessing the causal connection of any other parties or participants in the accident to the incident itself, and determining liability by levels of negligence, recklessness, or intent. Liability is assigned to the party determined to be the primary violator, or can be split among liable parties depending on their respective degree of fault.

The current framework of fault based liability can be better understood by considering examples – specifically, studying relevant case law. Hartley v. State[11], which came to a close in 1985, set a precedent for determining driver negligence and the

[11] Filed as a case against the state for negligently granting Johnson a driving license, but the direct causal link between Johnson's actions and the accident absolved the State of any liability.

subsequent liability attached. The case has over two thousand citing references, being invoked in 477 other cases. The outcome of Hartley v. State displayed a system that would assign liability, increasing the magnitude of consequences for repeat offenders. The background of the case is as follows, "Janet Hartley was killed when an automobile driven by Eugene R. Johnson crossed the center line and collided with her auto February 6, 1980, on a highway between Tacoma and Puyallup. Johnson was intoxicated at the time of the collision and was charged with negligent homicide."[12] The challenge facing the liability system in this case is determining the direction in which liability should move, as well as magnitude of the assigned liability. Although Johnson accepted liability costs with his guilty plea, there was little question of assigning liability elsewhere. He acted negligently by operating the vehicle while intoxicated and supplementary background information on Johnson showed a pattern of reckless behavior. Due to this background and the direct causal link between Johnson's actions and the damage caused, he was declared entirely liable for the accident.

A more recent case, Kazan v. Kennedy, which was decided in 2016, focuses on an accident with slightly different circumstances – involving a vehicle and a pedestrian. The case sites Hartley v. State, following its precedent. This case provides a view into an example of how the legal liability process unfolds in accident scenarios that don't involve just a single or multiple motorists. Kazan v. Kennedy displays the universality of the fault-based liability model, as it considers the causal connection to all patrons of the road and its fringes. The background of the case is as follows, "On the afternoon of July 23,

[12] Hartley v. State, 103 Wash. 2d 768, 770, 698 P.2d 77, 79 (1985)

2015, Plaintiff Samer Kazan, an Amazon employee, was walking south on Westlake Avenue in Seattle from one building in which he worked to another. Decl. of Cochran, Exh. 1 ("Kazan Depo") at 122:24-123:10, 178:9-10. Defendant Walter Kennedy (d/b/a Oncore Coach Leasing) was driving his bus north on Westlake Avenue. According to Plaintiff, the traffic light at the intersection of Westlake Avenue and Thomas Street showed a white "Walk" signal. Plaintiff testified that he entered the crosswalk and was struck by the bus driven by Defendant, who was turning left from Westlake onto Thomas."[13] Some witnesses corroborated the plaintiff's story, but only one did so with sufficient accuracy. While the witness account contributed to the verdict, the driver was declared to be negligently operating his vehicle for not noticing, and consequently yielding right of way to the pedestrian. Violation of the laws regarding right of way enabled the burden of liability to fall solely on the bus driver.

Although both examples represent driver liability, the precedent of causal connection between negligence and damage caused suggests that fault-based liability could have similarly found the pedestrian in Kazan v. Kennedy guilty and liable. It happened to be the case that the driver appeared to be more negligent than the pedestrian – who is declared free of claims of negligence, due to the light signaling that he had undisputed right of way. If instead the pedestrian had been distracted and stepped onto the crossing without noticing the light change to red, he would be considered the more negligent, and hence liable party. From these cases it seems that a similar process of judgement can be applied to a number of scenarios, as the purpose of liability is adhered

[13] Kazan v. Kennedy, No. C15-1251-BJR, 2016 WL 6084934, at *1 (W.D. Wash. Oct. 18, 2016)

to. From both cases it is clear that the guilty party was deemed as such due to violating laws and norms put in place to preserve a socially optimal state. Obstruction of the social optimum can be considered a moral flaw as it reduces the overall safety of road travel, increasing the risk for other vehicle users and pedestrians. The liability system applied in both cases condemns violations, discouraging future repetition.

The following sections discuss how fault based liability can still play a role in a new process of liability assignment – channeling our focus to forms of driver and manufacturer liability. The hypothesis here is that the current framework cannot cater to an AV world by itself, but aspects of it may still remain key to other models. The following sections will test this hypothesis.

Driver Liability

We begin our exploration of alternatives that we can acceptably employ. Although holding the manufacturer responsible seems the obvious second option, we will discuss other models before returning to it. A good starting point for our analysis would be the consideration of driver-directed liability. An adequate study of this type of liability requires answering questions of the moral permissibility of such a course of action; after all, a benefit of having an autonomous mode of transport is to allow more passenger independence by eliminating the need for constant control over the vehicle. The difficulty here comes in attempting to assign liability to drivers similar to the system for non-autonomous vehicles. Given the minimized degree of active driver intervention needed for AVs, the causal and agential link to driver actions is absent. Should methods of liability assignment differ if the driver has some opportunity for intervention; as opposed to none whatsoever? The analysis will follow the framework of Hevelke and Rumelin, dichotomizing driver liability into a 'duty to intervene' and 'strict liability' models. The former requires the driver to focus on road conditions enabling intervention when necessary; assigning liability to the driver in the event of failure to intervene. The latter model takes away both the duty and ability for intervention, but keeps in place the moral and legal responsibility for the accident, which can be construed as a supplemented financial responsibility.

Delving into the 'duty to intervene' model, it is apparent that the driver must pay continuous attention to the road, while also actively searching for risk and ultimately making decisions that appropriately address the risk at hand. Such requirements for the driver seem to match the basic criteria for vehicle operation that drivers currently faced with. Complete focus is presently a necessity due to the rate and potential of road accidents. Despite the imminence of road accidents, national statistics show that "at any given daylight moment across America, approximately 660,000 drivers are using cell phones or manipulating electronic devices while driving."[14] With the projected drastic decline in accident, fatality, and injury rates, it seems unlikely that an idle driver in an AV will exercise complete focus on the road and attempt to identify accidents that occur with significantly less probability.

[14] "Facts and Statistics." *Distraction.Gov*

The next question to answer is that of the driver's ability to predict potential accidents and intervene effectively. In fact, driver intervention could lead to less desirable outcomes than the vehicle's decision-making. Hevelke and Rumelin state:

> "Accidents are usually not easily foreseeable—especially if there is no driver that might be noticeably tired, angry or distracted. Therefore, it will probably be difficult to recognize dangerous situations which the autonomous vehicle might be ill equipped to manage, and even harder to intervene in time. Of course, much will depend on what kind of cases we are talking about. If the problem in which the driver must intervene tend to be foreseeable (if there is, for example, some sort of timely warning sign given by the vehicle), this is not a problem. But once we are talking about fully autonomous cars which drive as safely as the average person, such a predictability of dangerous situations seems unlikely and unrealistic. Moreover, accidents could not only happen because persons fail to override the system when they should have, but also because people override it when there really was no danger of the system causing an accident (Douma and Palodichuk 2012). As the level of sophistication of autonomous cars improves, the possibility of interventions by the driver might cause more accidents than it helps to avoid." (Hevelke, Rumelin 624)

Driver effectiveness is a significant uncertainty with dangerous potential, and seems to have a direct correlation with the likelihood of the occurrence of accidents. We have already established that the probability of accidents plummets with the introduction of

AVs. Since the likelihood of accidents is set to reduce in an AV world, it seems as though driver effectiveness will move in the same direction.

An additional concern that arises with a duty to intervene model considers the value of the AVs with regard to the added utility of such vehicles as compared to the current vehicles we are accustomed to. The concern of vehicle utility is as follows – a duty to intervene will eliminate the entire section of marginal vehicle users that AVs will gain access to. Marginal users include any individuals with an inability to drive. Adopting such a framework will negate the capability of independent vehicle operation granted by AVs by making it illegal for anyone without that very capability to use their vehicles. People with disabilities, and the elderly will be unable to enjoy the autonomy of the vehicles; and features that enable the vehicle to drive the owner home when drunk, or independently park itself will be rendered obsolete.

Presented with the combination of concerns of driver effectiveness and the lacking utility provided by this model, we are likely to see less enthusiasm from consumers of AVs. Potential AV users factor in the benefits of AVs into their choice to purchase such a vehicle – finding the time-saving ability to be less attentive, and the independence of self-sufficient transport greatly appealing. Additionally, it seems as though a Duty to Intervene model would be largely unacceptable to potential users of AVs, signaling the probable decline in consumer demand for the vehicles under such circumstances.

The ‘strict liability’ model pushes the burden of liability on the driver regardless of the inability to influence a potential accident situation. Even if the driver has no possible way of intervening, he takes on moral responsibility by understanding the risks that come with operating an AV. Although AVs are poised to significantly reduce road accidents, such events – and any associated risks – will continue to occur. The use of AVs will also continue to create risk for others, such as other vehicle operators, pedestrians, and even other passengers within the AV. Although blame cannot be assigned as a direct result of driver action, participation in the risky action of operating an AV, given the potential external costs, merits some burden of liability falling on the driver. The responsibility of risk is to be similarly assumed by all operators of AVs nationwide, a notion which dissects the strict liability model into two possible methods of implementation: (a) the current notions of strict liability in use (b) liability limited to risks associated with operating the AV.

The notion of strict liability is present very much present today, however is not invoked in the realm of auto accident liability. Strict liability can be seen in its application to pet ownership. Any pet owner assumes complete liability for damages caused by their pet, regardless of the circumstances surrounding the damage. For example, let’s take a person A who is in ownership of a dog who we’ll call Buddy. Buddy is an obedient and friendly dog with no history of ferocity. Regardless of this background, if Buddy does cause harm to someone, person A is the liable party; even in the event that Buddy was provoked and simply defending himself. Similarly, we can apply this framework to the ownership of AVs. The AV acts as a substitute for Buddy the dog, and liability will continue to fall on the owner, person A.

Even if no direct causal link to the driver can be made, liability must still be assigned, and may be levied as the overall cost of damages that resulted from the accident. It is likely that a majority of AV accidents will happen without driver cause, despite which drivers will need to bear the responsibility for any damage caused by their AVs. Avoiding liability becomes more a matter of probability and luck in such a situation.

The second method of implementation under the strict liability model accounts for the risks of AV operation as a collective, emphasizing the responsibilities to be undertaken daily by any user of an AV. The risks that come with AV use similarly affect every individual user, which suggests that the costs of the given risks must also be shared in a similar manner. A possible method of assigning liability costs involves observing liability at a macro level – collectively considering the parallel responsibilities of the entire AV ownership community. Despite freedom from direct causal blame, the potential risks and external costs tethered to using an AV can be assessed and immediately addressed preemptively by the institution of a mandatory ‘AV tax’. This is to say that in the event of accidents, AV drivers would have to bear no further liability costs as this burden would have been lifted prior. The tax enables AV drivers to atone for any potential risk, regardless of levels of negligence and causal relation in future accidents. The tax ensures that AV drivers maintain and account for their share of liability in all possible conditions.

Achieving strict liability in this manner requires in depth analysis of the various subparts that form the final product. Creating a mandatory AV tax comes with the convoluted task of determining the collective risks of the AV population, and further, arriving at a tax figure to levy on AV operators that will not discourage consumers from purchasing the vehicles. While the calculation of such a figure is well within the capability of federal powers, objections may arise during the process of distributing the costs. The structuring of the tax must be considered carefully to avoid inequities to the greatest extent possible – it is possible that numerous AV users will be paying costs that are surplus to their road safety patterns, and thus are contributing in excess of the just and equitable amount assigned by societal factors. On the flipside, a number of AV users are likely to contribute less than their fair share to the societal costs.

Further, strict liability models seem to have some moral flaws; as it can unjustly distribute punishment. The Kazan v. Kennedy case can be modified to provide an example – substituting the bus driver with an AV user. Suppose in this scenario that the walkway signal had turned red, but the pedestrian, distracted on his cell phone, entered the walkway unaware. The AV user making the turn had the right of way, but the vehicle collides with the crossing pedestrian. Even if the AV user was exerting the utmost care, and the vehicle was operating as per programming and guidelines, the AV user is the liable party. It appears in this case that it would be morally wrong to implicate the AV user in lieu of the pedestrian. The assigning of liability to the AV user could conflict with ideas introduced in the section discussing the purpose of liability.

An aspect of the purpose of liability encourages moral improvement to preserve an acceptable level of morality that is required to prevent and minimize accidents. By assigning liability without fault and causation, we are hindering the ability of individuals to exercise their rights by punishing[15] them for liability, despite an inability to avoid it. It appears that in our pursuit of an ideal social level of morality, we are committing moral violations. "Any form of strict liability that permits criminal conviction of defendants that lack moral fault is unacceptable (Simester 21, Gardner 69–70, Husak 91, Duff 128). The moral framework that supports this claim is rights-based (Simester 34), relying on the connected thoughts that criminal conviction leads to punishment, punishment must be deserved, desert implies wrongdoing, and wrongdoing presupposes fault."[16] Although there is a mention of criminal conviction, the argument can extend to any methods that impose a restraint on human autonomy and rights. Punishment for criminal convictions could take the form of monetary fines, forfeiture of property, imprisonment and death, among others. These forms of punishment place a constraint on the offender's freedom either by limiting resources or physically limiting the ability to act in a manner of choice. Similarly, the consequences (whether they be fines, community service, imprisonment etc.) of being involved in an accident are applied as a constraint on the AV user. Strict liability forces AV users to accept constraints on their autonomy and rights even if they acted perfectly in accordance with guidelines.

---

[15] We can speculate that punishment is likely to be a financial burden; perhaps along with some penalty on driving credibility similar to the three strike license system today. It is possible that imprisonment may serve as an option, but seems improbable given the forced nature of strict liability.

[16] Shiner, R.A. & Hoemsen, J. Criminal Law, Philosophy (2007) 1: 119. doi:10.1007/s11572-006-9007-9

The expectation is that AVs will be manufactured with operational thresholds that match consumer demand, while also maximally diminishing risk. Individuals will likely refuse payment for any potential AV accident in the taxed region due to the subjectivity and difficulty in accurately quantifying the potential dangers of operating an AV. It seems as though this model of driver liability will also be deemed socially unacceptable.

Additionally, some manufacturers (Mercedes, Volvo, and Google as of now) have publicly announced their intention to bear the complete burden of liability, should their vehicles be involved in any accident. Drivers of these vehicles will no longer be held liable, thus this decision by manufactures is likely to push us away from models of driver liability[17].

We are now left with the analysis of this manufacturer based model of liability – one in which all manufacturers bear the full burden of accident liability. The section below discusses the possibilities that come with such a process, as well as different cases in which liability processes will have to be varyingly applied.

Manufacturer Liability

The previous section effectively rules out models of driver liability, leaving us with another unexplored option – if we can't assign liability to the people involved focus must shift to the other victim of any accident – the vehicle. Such personification of the

[17] There remains a possibility of driver liability maintain relevance – in the event that not all AV manufacturers follow this trend. If some manufacturers choose not to take on liability we are once again faced with a discussion of how to assign liability: should driver liability models be applied to for these AVs, or should we force manufacturer liability on all manufacturers?

machine may seem striking, but holds significant weight and relevance in the discussion to follow. The functioning of the AV involves the mimicking of certain human traits and behaviors – specifically the assessment and addressing of imminent risks. Despite having an immensely higher threshold for error as compared to humans, the vehicles don't escape the set of external costs that come with their operation. Given the prospect of damage (regardless of an overall decrease in potential damages), an important question arises: should the burden of liability be assigned based on shortcomings of the vehicles? After all the vehicle is the principal decision-maker. AVs independently observe, calculate, assess, and determine methods of addressing on-road risk; ultimately carrying out the course of action that caused damage. There remains the chance that a human driver would react to presented stimuli in a manner that results in less collateral damage, hence further incriminating the flaws of the vehicle. It can at least be said with certainty that the consumer will benefit from this system, being distanced from concerns of liability.

Every purchaser and user of AVs will be acquiescent to this, as the burden of liability is entirely lifted and passed on to the vehicle. It is important to keep in mind that the rate of AV accidents is not being altered, thus maintaining the moral value of this machine, and affording users the additional benefit of lower-risk commuting.

Assigning liability to the vehicle provokes the question of how the vehicle itself can address this liability. We have followed a causal chain here that skipped past drivers and has linked to the vehicle. Consideration of the purpose of liability needs to be invoked in this situation. Based on our prior discussion of this purpose, it seems that we cannot achieve the aims of liability if it is applied solely to the vehicle; because despite its autonomous decision making capabilities, an AV has no moral or fiscal agency that independently incentivizes the vehicle to act in accordance with liability systems. Thus, we have to trace the causal link further back to determine an adequate agent. The decision-making capabilities of any AV are based on a set of moral guidelines provided during the production process, thus indicating that any agency the vehicle has is an extension of that of its manufacturer. Reapplying our thinking of the purpose of liability to the agent – in this case the manufacturer – informs us that we have found a sufficient target for liability.

Prior to delving into various forms of manufacturer liability, as well as manufacturers' mindset regarding liability adoption, it is helpful to discuss a base case – situations in which the manufacturer can be indubitably held liable for accidents: establishing a causal link to the manufacturer. This requires taking a step back to our analysis of the current framework of liability assignment. Fault reemerges here as a determinant of liability, once again using themes of negligence, recklessness, and intentional misconduct as axes of assessment. Using such a model of fault liability channels focus to any errors throughout the process of manufacture. The process can be broadly divided into three subparts – vehicle design, production or marketing. Any errors committed by people employed in these positions establishes the manufacturer's causality[18]. The identification of any flaws in these three areas that would increase the probability of an accident justify liability being applied to the manufacturer. The

[18] This will be the case unless an employee is acting under their own prerogative, outside of the manufacturer's extended agency. This notion is briefly mentioned further down.

determination of the magnitude of the liability penalty will be determined on categorizing flaws according to the tiers of fault. The definition and consequence of fault will be taken from the existing federal guidelines – the construction of which is beyond the scope of this paper. With this understanding of the condemnation of incriminating flaws during manufacture we can consider possible examples.

Potential defects in the car can occur from flaws in the design process. Simple flaws in vehicle design can have a detrimental impact in the event of an accident. For example, the placing, as well as the number of sensors feeding the vehicle's decision-making system must be carefully analyzed. Further, design features may prevent the vehicle from operating under certain circumstances – the aerodynamics of the vehicle may disallow it from travelling beyond certain speeds, and certain factors may even declare it dangerous to operate in specific climates or road conditions. Marketing the vehicles in regions that exhibit conditions outside of the vehicle's operational capabilities can constitute negligence on part of the manufacturer. Similarly, it is the responsibility to provide and make evident and accessible any cautionary literature that informs users of potential hazards. Any such feature of the vehicle that conflicts with optimal safety measures can amount to manufacturer fault, assigning liability entirely to the manufacturer, and impeding progress by damaging the manufacturer’s reputation and financial well-being.

Manufacturers can also be causally linked to accidents through the code that enables autonomy in the vehicles. We look back to the earlier discussion on the code of ethics programmed into the vehicle. Since the vehicle’s agency is an extension of the manufacturer's, it is imperative that the code of ethics implanted in the vehicle is morally sound. Any moral shortcomings will render the manufacturer liable. Agency can be further broken down to assign liability. It is possible that employees are acting under their own agency, in which case the system is likely to hold the responsible individuals liable. Liability will remain with the manufacturer if employees are acting as agents of their employers.

Focusing on the latter – manufacturer liability – we continue to explore an ideal framework. From a market standpoint, it is likely that consumers will understand the benefits they stand to gain from manufacturer liability and flock to purchase AVs – increasing the overall market demand. Although the demand for AVs will rise, it may be met by a suffocating supply due to the heightened risks for manufacturers of AVs. Manufacturers may be unable to accurately predict the number of accidents that occur; and negative economic implications could come with a prediction that falls short of the true cost. Any resulting constraints present a challenge to this simple version of manufacturer liability that this paper will call the Hindrance problem. The substance of the problem is that manufacturers will have less incentive to produce AVs due to the high risk of liability and litigation costs.

> "The technology is potentially doomed if there are a significant number of such cases, because the liability burden on the manufacturer may be prohibitive of further development. Thus, even though an autonomous vehicle may be safer overall than a conventional vehicle, it will shift the responsibility for accidents, and hence liability, from drivers to manufacturers. The shift will push the

> manufacturer away from the socially-optimal outcome—to develop the autonomous vehicle" (Marchant and Lindor 1334)

An important notion presented in this passage is that of a socially optimal outcome. Optimality seems to be tethered to the development of AVs – and the threat of liability costs could discourage manufacturers from investing in further development of the technology, or abandoning production altogether. Branching out from Marchant and Lindor's worry of abandoning production, we can speculate another possibility – that the burden of liability will discourage manufacturers from pursuing socially optimal vehicle behaviors in favor of behaviors that achieve outcomes with the lowest potential costs to the manufacturer. Liability costs may be significant enough to damage the financial well-being of manufacturers, suggesting that in the event that liability costs cannot be sufficiently reduced, manufacturers may decide to abandon the production of AVs altogether. This outcome is socially undesirable, as it foregoes the numerous additional lives that can be saved, and damage that can be prevented. The hindrance problem seems to be a significant hurdle for this model of manufacturer liability, hence our new framework must directly address this issue.

The solution needs to provide incentive to manufacturers to continue the production of AVs, and simultaneously ensure that the purpose of liability is fulfilled. However, not all methods that promote deterrence and rehabilitation will satisfy manufacturers. An example of this is assigning a tax to all manufacturers in a similar manner to the strict liability model under driver liability. While this achieves the purpose of liability, it does not address the hindrance problem. Perhaps incapacitation can add value to this liability system.

In light of this notion of incapacitation, Hallevy lends us another hand here with the introduction of his Corporate Imprisonment model for manufacturer liability. Let us first consider the semantics of the term. Imprisonment can be understood simply as a deprivation of liberty, in this case applied to corporations producing AVs. We can immediately rule out certain forms of incapacitation such as imprisonment and sanctions on the manufacturing process, as it will choke the production of AVs and impede our social progress toward optimal social good. We now seek a different form of liberty deprivation – one which works toward maintaining the market level – production and development of AVs.

A possible method of solving this dilemma is the siphoning *and* sharing of resources. The primary shortcoming of a tax penalty model was the diminished market output of AVs, and the lacking incentive for manufacturers to maintain production to standards. Instead of gathering the excess tax revenue, the taxed amount can be shared with a competitor in the form of a subsidy. Viewing this transaction from an industry-wide perspective, the changes balance out leaving the overall fiscal state of the industry the same; although the capabilities of the manufacturers involved in the transaction are altered.

Such a tax/subsidy model would not only incentivize liable manufacturers to prevent future incidents; and to do so efficiently given the advantaged position of a competitor. The competitor that is rewarded with a subsidy must be determined by a merit-based system to further incentivize competition. The features and functioning of this system lays beyond the scope of our analysis.

The likely scenario for AV liability is one in which manufacturers agree to take on the burden of accident liability. Although manufacturers may pass on some of the liability costs to consumers through higher prices and servicing costs, the manufacturer remains the directly liable agent. Accepting liability urges manufacturers to develop and produce the optimal vehicle to better prevent accidents and the ensuing liability process. It seems that such a model achieves the purpose of liability considering manufacturers, but we must ensure that the same occurs with other agents. AV users may exploit the lack of liability and behave in a reckless manner – taking action such as operating the vehicle in unadvised conditions. The liability framework must act as a deterrent to users as well as manufacturers. To prevent user abuse, the fault based aspects of the current framework of liability can be invoked. Although manufacturers are responsible for the liability costs, they can be granted an exemption if an accident is caused by user negligence or recklessness.

IV. Conclusion

As we move toward an autonomous world the need for user input diminishes, signaling the need for expansion in our user-centric legislation. Transportation will be based on the decision making of an AI system; thus the adopted liability framework must take into account the extended capabilities of vehicles. The framework must also set a precedent for accountability and discourage accident inducing behavior. The emphasis of both the purpose of liability and the willingness to purchase or produce AVs is important, as the framework must aim to jointly preserve both features.

Possible liability frameworks were assessed to determine the adherence to this required balance. Considering both models of driver liability, it seems likely that consumer willingness for AVs will decline. These models present moral flaws that a just system must avoid. Additionally, the purpose of liability is not universally applied to all parties, as manufacturers have the ability to exploit a system which places all liability on vehicle users.

Similarly, models of manufacturer liability face the inverse problem of the purpose of liability not being entirely fulfilled, as AV users have the potential to exploit the lack of punishment. Manufacturers too would be more accepting of a framework that understands potential liability concerns beyond the manufacturer's control that could harm the resources needed for AV production and development. Manufacturer liability is likely to become the instituted liability framework, but needs to address the potential injustices it allows.

Although the current framework of liability assignment needs expansion, it appears that the causal aspects of the model can fill the gaps in the other liability models considered. Manufacturer liability is likely to become the dominant framework of liability in an autonomous world; but in order for this framework to fit an ideal mold, it must incorporate an analysis to determine fault. Causal links remain of immense importance, and are a direct determinant of liability, which accentuates the value of aspects of the current framework of liability.

References

- Ackerman, Evan “People Want Driverless Cars With Utilitarian Ethics, Unless They’re a Passenger.” *IEEE Spectrum,* 23 Jun. 2016, http://spectrum.ieee.org/cars-that-think/transportation/self-driving/people-want-driverless-cars-with-utilitarian-ethics-unless-theyre-a-passenger. Accessed 8 Nov 2016.

- “National Statistics.” *Fatality Analysis Reporting System (FARS) – National Highway Traffic Safety Administration (NHTSA),* 2015, http://www-fars.nhtsa.dot.gov/Main/index.aspx. Accessed 6 Nov 2016.

- “Traffic Safety Facts – Crash Stats.” *U.S. Department of Transportation – NHTSA*, Nov 2015, https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812219. Accessed Nov 6 2016.

- Urmson, Chris. “The View from the Front Seat of the Google Self-Driving Car.” *Backchannel*, 11 May. 2015, https://backchannel.com/the-view-from-the-front-seat-of-the-google-self-driving-car-46fc9f3e6088#.u4fkn3416. Accessed 7 Nov 2016.

- “Fault and Liability for Motor Vehicle Accidents.” *FindLaw*, http://injury.findlaw.com/car-accidents/fault-and-liability-for-motor-vehicle-accidents.html. Accessed 22 Nov 2016.

- “Facts and Statistics.” *Distraction.Gov*, 2016, http://www.distraction.gov/stats-research-laws/facts-and-statistics.html. Accessed 20 Nov 2016.

- Shiner, R.A. & Hoemsen, J. *A.P Simester (ed): Appraising Strict* Liability. Criminal Law, Philosophy (2007) 1: 119. doi:10.1007/s11572-006-9007-9

- Hallevy, Gabriel. When Robots Kill : Artificial Intelligence under Criminal Law. Boston, US: Northeastern, 2013. ProQuest ebrary. Web. 30 November 2016.

- Hevelke, A. & Nida-Rümelin, J. *Responsibility for Crashes of Autonomous Vehicles: An Ethical Analysis*. Sci Eng Ethics (2015) 21: 619. doi:10.1007/s11948-014-9565-5. http://link.springer.com.ccl.idm.oclc.org/article/10.1007%2Fs11948-014-9565-5. Accessed 25 Nov 2016.

- Tonkens, R. *A Challenge for Machine Ethics.* Minds & Machines (2009) 19: 421. doi:10.1007/s11023-009-9159-1. http://link.springer.com.ccl.idm.oclc.org/article/10.1007%2Fs11023-009-9159-1. Accessed 21 Nov 2016,

- Anderson, M. & Anderson, S.L. *The Status of Machine Ethics: A Report from the AAAI Symposium.* Minds & Machines (2007) 17: 1. doi:10.1007/s11023-007-9053-7.

http://link.springer.com.ccl.idm.oclc.org/article/10.1007%2Fs11023-007-9053-7. Accessed 20 Nov 2016.

- Gunderson, J.P, Gunderson L.F. “Autonomy (What’s it Good for?).” *ACM Digital Library,* 30 Aug. 2007, http://delivery.acm.org/10.1145/1670000/1660896/p141-gunderson.pdf?ip=134.173.173.72&id=1660896&acc=ACTIVE%20SERVICE&key=2E7FD48FA6D30C27%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&CFID=698929482&CFTOKEN=32125856&__acm__=1480561249_34e99bdda73d7ebc4f2799cdf581e101. Accessed Nov 24 2016.

- Bonnefor, J.F, Shariff, A, Rahwan, I. “The Social Dilemma of Autonomous Vehicles.” *Science*, Vol. 352, Issue 6293, pp. 1573-1576, 24 Jun. 2016, http://science.sciencemag.org/content/352/6293/1573.full. Accessed 17 Nov 2016.