ETHICS AND PUBLIC HEALTH OF DRIVERLESS VEHICLE COLLISION PROGRAMMING

SAMANTHA GODWIN*

Driverless vehicles present a core ethical dilemma: there is a public health necessity and moral imperative to encourage the widespread adoption of driverless vehicles once they become demonstrably more reliable than human drivers, given their potential to dramatically reduce automobile fatalities, increase autonomy for disabled people, and improve land use and commutes. However, the very technologies that could enable autonomous vehicles to drive more safely than human drivers also imply greater moral responsibility for adverse outcomes. While human drivers must make split-second decisions in automobile collision scenarios, driverless car programmers have the luxury of time to reflect and choose deliberately how their vehicles should behave in collision scenarios. This implies greater responsibility and culpability, as well as the potential for greater scrutiny and regulation. Programmers must make premeditated decisions regarding whose safety to prioritize in inevitable collision

* Resident Fellow, Yale Law School, Information Society Project. I am very grateful to Jack Balkin, Lea Brilmayer, Rebecca Crootof, Kay Dannenmaier, Erik Duhaime, Daniel Hornal, Mason Marks, Clare Ryan, Rebecca Wexler, Gideon Yaffe, and the participants in the Information Society Project Resident Fellowship Workshop for their helpful comments on this article.

scenarios—situations where a vehicle cannot avoid a collision altogether but can choose between colliding into different vehicles, objects, or persons.

With the recent bipartisan passage of the SELF DRIVE Act in the House and the rapid development of driverless vehicle technology, we are now entering a critical time frame for considering what priorities should govern driverless car inevitable collision behavior. This Article shall argue that prescribed "ethics" programing must be regulated by law in order to avoid the likely collective action problem of a marketplace that will reward "occupant-favoring" designs, despite a probable public preference (and arguable moral necessity) for occupant indifferent designs. This Article then considers a variety of different options for systems of driverless vehicle ethics programming. The most justifiable ethics programing system would be one where road users are discouraged from externalizing the dangers incurred by their transportation choices onto those whose transportation choices, if more widely adopted, would comparatively improve aggregate safety. This ethical programing system, which I term "incentive-weighted programing," would promote public safety while also striking the most equitably justifiable balance between different road users' interests.

INTRODUCTION

Widespread adoption of self-driving[1] cars capable of nearly error-free driving could potentially save tens of thousands of lives every year. More than 37,000 people died in automobile collisions in the United States in 2016,[2] and approximately 1.2 million people are killed in automobile accidents around the world each year, according to the World Health Organization.[3] Driver error is believed to account

[1] Throughout this Article I will use the terms "driverless cars," "autonomous vehicles," "driverless vehicles," or "self-driving cars" to refer to automobiles that steer, brake, and accelerate without contemporaneous input from their occupants or human operators. I will use the terms "manual cars" or "manual vehicles" to describe vehicles that are primarily steered by a human driver regardless of elements of automation such as parking assistance, cruise control, or automated accident avoidance. While Tesla Autopilot is often discussed in the context of driverless vehicles, this system does not constitute a "driverless vehicle" in the sense considered in this Article since Tesla permits and relies on human steering control in hazardous conditions.

[2] U.S. DEP'T OF TRANSP., NAT'L HIGHWAY TRAFFIC SAFETY ADMIN., QUICK FACTS 2016 1 (2017), https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812451. The year 2016 is the most recent year for which the U.S. Department of Transportation published fatal collision numbers. *See id.*

[3] WORLD HEALTH ORG., GLOBAL STATUS REPORT ON ROAD SAFETY 2015 vii (2015), http://www.who.int/violence_injury_prevention/road_safety_status/2015/en/.

for more than 90% of automobile collisions.[4] These facts taken together suggest that the widespread adoption of driverless vehicles that navigate roads more reliably than human drivers would provide immense public health benefits, potentially saving enormous numbers of lives.

Autonomous vehicle technology may already be close to surpassing the safety of typical human drivers. As of March 2018, driverless vehicles were believed to have only been implicated in a single known fatal collision,[5] an incident where local police believed the accident was not due to a fault of the car and would have been difficult to avoid.[6] Google's current self-driving car prototype has been driven over 1.3 million miles since 2009.[7] As of February 2016, they have caused only one collision.[8] This driving record clearly does not provide a complete picture of their capabilities if adopted as widespread replacements for manually driven vehicles, since Google's driverless cars are supervised by human drivers able to intervene in dangerous situations[9] and are restricted to safe weather conditions and well-mapped roads.[10] However, as the technology progresses

[4] *See* Bryant Walker Smith, *Human Error as a Cause of Vehicle Crashes*, CTR. FOR INTERNET & SOC'Y (Dec. 18, 2013, 3:15 PM), http://cyberlaw.stanford.edu/blog/2013/12/human-error-cause-vehicle-crashes (reviewing various studies that place the rate of human error causing or contributing to crashes as between 90% and 99%).

[5] *See* Daisuke Wakabayashi, *Woman's Death in Arizona Casts a Pall on Driverless Car Testing*, N.Y. TIMES, Mar. 20, 2018, at A1.

[6] *See* Carolyn Said, *Exclusive: Tempe Police Chief Says Early Probe Finds No Fault by Uber*, S.F. CHRON., (Mar. 26, 2018), https://www.sfchronicle.com/business/article/Exclusive-Tempe-police-chief-says-early-probe-12765481.php. A motorist in a Tesla on autopilot mode was also killed in a collision with a truck. *See* Danny Yadron & Dan Tynan, *Tesla Driver Dies in First Fatal Crash While Using Autopilot Mode*, GUARDIAN (June 30, 2016), https://www.theguardian.com/technology/2016/jun/30/tesla-autopilot-death-self-driving-car-elon-musk. The Tesla system, however, is not intended to be fully autonomous. *See id.* Rather, it requires driver attention and even so, has a safety record that outperforms manually driven vehicles with the first death in 130 million miles of Tesla autopilot driving as compared to a death every 94 million miles driven manually. *Id.*

[7] *See* Alex Davies, *Google's Self-Driving Car Caused Its First Crash*, WIRED (Feb. 29, 2016, 2:04 PM), https://www.wired.com/2016/02/googles-self-driving-car-may-caused-first-crash/.

[8] *See id.*

[9] *See* THIERRY FRAICHARD, WILL THE DRIVER SEAT EVER BE EMPTY? 3 (2014), https://hal.inria.fr/file/index/docid/968002/filename/14-rr-fraichard.pdf.

[10] *See* Lee Gomes, *Hidden Obstacles for Google's Self-Driving Cars: Impressive Progress Hides Major Limitations of Google's Quest for Automated Driving*, MIT TECH. REV. (Aug. 28, 2014), http://www.technologyreview.com/news/530276/hidden-obstacles-for-googles-self-driving-cars/ (noting that Google self-driving cars rely on pre-mapped roads and have not yet operated in the snow and rain).

further, we can be increasingly confident that self-driving vehicles will eventually surpass the reliability and precision of typical motorists.[11]

Beyond the potential safety advantages of driverless vehicles, driverless cars have a tremendous potential to immensely improve the independence of people with disabilities, make commutes less stressful, reduce congestion, and enable more efficient use of dense urban spaces.[12] Without the need for the occupant to have any of the sensory or motor skills needed to operate a conventional automobile, driverless cars would enable disabled and elderly people greater personal independence and self-reliance.[13] Driverless vehicles could park themselves in places people would find inconvenient after dropping off passengers, reducing the need for urban parking lots.[14] Driverless car sharing services could allow one vehicle to serve the needs of many people without requiring a human driver. They would also allow people who are unable to drive safely to have the freedom to travel on roads independently, without the need for licensing tests.[15]

Driverless vehicle designers and regulators can be expected to produce driverless cars that aim to avoid collisions whenever possible and practicable.[16] Not all automobile collisions, however, are due to

[11] *See* Tom Simonite, *Data Shows Google's Robot Cars are Smoother, Safer Drivers Than You or I: Tests of Google's Autonomous Vehicles in California and Nevada Suggests They Already Outperform Human Drivers*, MIT TECH. REV. (Oct. 25, 2013), http://www.technologyreview.com/news/520746/data-shows-googles-robot-cars-are-smoother-safer-drivers-than-you-or-i/.

[12] *See* Sam Lubell, *Here's How Self-Driving Cars Will Transform Your City*, WIRED (Oct. 21, 2016), https://www.wired.com/2016/10/heres-self-driving-cars-will-transform-city/.

[13] *See, e.g.*, SELF DRIVE Act, H.R. 3388, 115th Cong. § 9(b)(e)(1) (as passed by House, Sept. 7, 2017).

[14] *See* Lubell, *supra* note 12.

[15] *See* Henry Claypool, *Self-Driving Cars: The Impact on People with Disabilities*, RUDERMAN FAM. FOUND. (Jan. 2017), http://rudermanfoundation.org/wp-content/uploads/2017/08/Self-Driving-Cars-The-Impact-on-People-with-Disabilities_FINAL.pdf.

[16] Accidents have costs, but there are also secondary and tertiary costs to reducing "primary accident costs," since reducing accidents often requires forgoing other opportunities for cost reduction, or creating costly incentives, such that reducing the risks of accidents beyond certain points can sometimes generate more risks overall. For a discussion of these issues, see GUIDO CALABRESI, THE COST OF ACCIDENTS: A LEGAL AND ECONOMIC ANALYSIS 29, 131–236 (1970). For example, bicycle helmet laws are controversial given the possibility that they may discourage cycling by making it less convenient. Nudging would-be cyclists towards less health-promoting or more dangerous forms of transportation could exceed the health benefits found in increasing the portion of cyclists who wear helmets. *See, e.g.*, Oliver Milman, *Mandatory Bike*

potentially correctable driver error; some are due to road and environmental hazards, mechanical failure, or pedestrian or animal crossings.[17] No matter how well driverless vehicles are designed, they will likely encounter "inevitable collision states" where a collision can be anticipated but not avoided. The concept of "inevitable collision states" was developed by Thierry Fraichard in his report, *Will the Driver Seat Ever Be Empty?*, where he presents a case that, given real world environmental conditions, it is physically impossible to design a driverless vehicle that can be guaranteed to avoid all collisions.[18] When some collisions are unavoidable, driverless vehicle programming, like human drivers, will often have to select what object or person they collide with.[19] For example, if a group of deer dart in front of a quickly moving car, the driver may have to choose between colliding with the deer, swerving into oncoming traffic, or swerving off the road and colliding with a tree. The option to avoid hitting anything may be unavailable given the speed of the car and surrounding obstacles.

In real world cases, there may be relatively few *fully* inevitable collision states where *some* collision can be entirely predicted regardless of a vehicle's steering and braking. There are likely to be many more cases where technical or practical uncertainties present scenarios where there is a clear *risk* of a collision that cannot be completely eliminated but no collision is certain. In these cases, however, there are still necessary choices that affect which object stands the greatest risk of being struck. In the previous example of a group of deer darting in front of a quickly moving car, a human motorist may choose among braking without swerving, increasing the relative risk of colliding with the deer, swerving left, increasing the relative risk of colliding with oncoming traffic, or swerving right,

Helmet Laws Do More Harm Than Good, Senate Hears, GUARDIAN (Aug. 12, 2015), https://www.theguardian.com/lifeandstyle/2015/aug/12/mandatory-bike-helmet-laws-do-more-harm-than-good-senate-hears.

[17] These and other causes are reviewed and discussed extensively in NAT'L HIGHWAY TRAFFIC SAFETY ADMIN., U.S. DEP'T OF TRANSP., NATIONAL MOTOR VEHICLE CRASH CAUSATION SURVEY, REPORT TO CONGRESS 2, 23, 25 (2008), http://www-nrd.nhtsa.dot.gov/Pubs/811059.pdf.

[18] FRAICHARD, *supra* note 9, at 12; *see also* Thierry Fraichard & Hajime Asama, *Inevitable Collision States: A Step Towards Safer Robots?*, PROC. 2003 IEE/RSJ INTL. CONF. ON INTELLIGENT ROBOTS & SYS., Oct. 2003, at 388, https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1250659 (Fraichard's and Asama's original definition of the concept of inevitable collision states).

[19] Patrick Lin, *The Ethics of Autonomous Cars*, ATLANTIC (Oct. 8, 2013), http://www.theatlantic.com/technology/archive/2013/10/the-ethics-of-autonomous-cars/280360/.

increasing the relative risk of colliding with a tree. The human motorist would presumably try to execute each of these possible maneuvers in a way that minimized the probability of any collision, but each option predictably increases the risk of colliding with one object relative to the risk of colliding with another object.

There are some clear-cut cases where it is better to strike one object than to strike another when avoiding a collision altogether is impossible. For example, in a scenario where a vehicle is being forced out of its lane and could collide with either a water-filled traffic barrier[20] or into a boulder, it would obviously be preferable for the car to strike the barrier and not the boulder to minimize damage to the vehicle and its occupants.[21]

In collision scenarios involving other people and their property, however, the choice of collision behavior programming will be a choice of how to allocate risk between people and whose safety to prioritize.[22] This poses significant ethical dilemmas for how driverless vehicles should allocate risk in likely or inevitable collision states, and who should decide how those risks are allocated. In some collision scenarios, the choice of one programing option over another will determine which passengers or pedestrians survive the crash.

Congress has just begun the task of considering driverless car regulation through the bipartisan passage of H.R. 3388, the SELF DRIVE Act, in the House in late 2017.[23] The SELF DRIVE Act does not address collision programing (often called "ethics programing"[24]) directly; instead, the Act establishes a Highly Automated Vehicle

[20] Water-filled barriers or sand-filled barriers, sometimes termed "impact attenuators," are frequently used as a safer means of controlling the flow of traffic than solid barriers, in part because automobile collisions with such barriers are likely to do less damage to the car and those around it than solid barriers. *See Frequently Asked Questions: Barriers, Terminals, Transitions, Attenuators, and Bridge Railings*, U.S. DEP'T TRANSP., FED. HIGHWAY ADMIN. (Aug. 31, 2017), https://safety.fhwa.dot.gov/roadway_dept/countermeasures/faqs/qa_bttabr.cfm; *see also Crash Cushions*, U.S. DEP'T TRANSP., FED. HIGHWAY ADMIN. (Nov. 2013), https://safety.fhwa.dot.gov/roadway_dept/countermeasures/docs/CrashCushions_Nov2013Safelogo.pdf.

[21] *See* Lin, *supra* note 19.

[22] The question of priority does not arise in cases where a collision can be avoided altogether or is unavoidable but a driverless car has no ability to affect how the damage is allocated.

[23] SELF DRIVE Act, H.R. 3388, 115th Cong. § 9(b)(e)(1) (as passed by House, Sept. 7, 2017).

[24] *See* Jamie Carter, *Automated Cars and AI: Reasons Why the Tech Industry Must Consider Ethics*, TECHRADAR (Mar. 13, 2015), http://www.techradar.com/us/news/world-of-tech/automated-cars-and-ai-reasons-why-the-tech-industry-must-consider-ethics-1287455.

Advisory Council of the National Highway Traffic Safety Administration to study driverless vehicles and issue a report to Congress for possible future regulation.[25] We are now, therefore, entering a critical time frame to seriously reflect on and debate appropriate driverless vehicle regulation prior to their widespread consumer use.

The widespread adoption of reliable driverless cars would provide extraordinary benefits while also creating several ethical dilemmas. Who should determine what principles guide driverless vehicle ethics programming, and what should those principles be? This Article shall argue that the objectives of ethics programing must be regulated by law in order to overcome a probable collective action problem: in a marketplace where ethics programming is unregulated, strongly "occupant favoring" designs will be rewarded despite a probable public preference (and moral necessity) for occupant indifferent designs.[26] There is also a tension between the moral imperative to encourage the use of reliable driverless vehicles given their potential to reduce fatal collisions, with the enhanced moral responsibility for the distribution of harm in collisions they cannot avoid, given that their behavior is preprogrammed. Since driverless vehicle behavior in collisions at least in part reflects a premeditated decision by their programmers[27] rather than the instinctual reaction of a driver, the choice of whose safety to prioritize and why requires greater ethical scrutiny, and unjustifiable prioritizations imply greater moral culpability. This Article argues that the most ethically justifiable and reasonable ethics programing system is one where road users are discouraged from externalizing the hazards of their transportation choices onto others who adopt modes of transportation that pose less risk to others in the aggregate. This ethical programing system, which I term "incentive-weighted programing," provides a novel and defensible account of how to factor-in the morally salient, distinctive characteristics of driverless cars in settings where they coexist with pedestrians and manually driven vehicles.

[25] SELF DRIVE Act, H.R. 3388, 115th Cong. § 9(b)(e)(1) (as passed by House, Sept. 7, 2017).

[26] An "occupant favoring" design prioritizes the safety of the vehicle's occupant over the safety of other people, for no reason other than they are occupying the vehicle in question. An occupant indifferent design does not make this prioritization choice, though there may be other grounds on which it favors the safety of the occupant.

[27] Driverless vehicles might be programmed in ways that make few, if any, of their collision behavior "choices" predictable ex ante by their programmers, but this would also be a choice with risk distributive consequences.

In Part I of this Article, I will identify some of the ethical dilemmas that will need to be addressed in the programming of driverless vehicle collision behavior. The design of driverless vehicle behavior in inevitable collision states present real world "trolley problems"[28] that prompt unavoidable ethical choices. In Part II, I will address the question of who should decide the ethics programming of driverless vehicles. I will make a case for the necessity of regulating uniform collision behavior programming prior to the widespread adoption of driverless vehicles, rather than leaving these programming decisions up to customer or manufacturer choice. This is because, absent regulation, a collective action problem is likely to occur where companies are incentivized to adopt "occupant favoring" collision programing designs, but the public at large would most likely prefer non-occupant favoring designs. Given background conditions where a vehicle may be programmed to be either occupant-favoring or occupant-indifferent, individual consumers would be safer choosing occupant favoring programming for themselves. As between a rule of general applicability that driverless vehicles must be occupant-indifferent or must be occupant-favoring, however, the public would be safer with an occupant-indifferent mandate. This suggests that safety-motivated consumers would choose a different programming option to govern all driverless vehicles then the option the market will deliver if left up to individual choice. In Part III, I will address the question of what principles of ethics programming ought to be adopted for driverless cars. I will outline and evaluate different options for ethics programming, including systems that reflect different utilitarian designs, fault-based designs, Coasian considerations, and designs that regard the vehicle as having a duty of loyalty to its user. I will conclude that regulators should adopt an "incentive-weighted" system, rather than a utilitarian system, occupant-favoring system, or fault-based system.

[28] *See* Philippa Foot, *The Problem of Abortion and the Doctrine of the Double Effect*, *in* VIRTUES AND VICES AND OTHER ESSAYS IN MORAL PHILOSOPHY 19 (1978) (the original account of the Trolley Problem).

I. Outlining the Dilemmas

When human drivers face scenarios that force them to choose between colliding with one object or another, they have virtually no time for reflection or deliberation.[29] This mitigates their moral culpability for poor or selfish decisions, since their decisions are made without premeditation or time for consideration. Human motorists are legally required to use reasonable care such that they avoid acting negligently.[30] They are not, however, required or expected to behave optimally, and their limits are factored into determining standards for "reasonable care." Human drivers are responsible for most automobile accidents,[31] but they can often be excused legally and morally, given human limitations in reaction time and judgment under the split-second decisional conditions of automobile accidents.[32] Even when human motorists' driving behavior is inexcusable, accidents are typically the result of negligence or recklessness, where they would have done otherwise if they could revise their driving behavior.

Driverless vehicle behavior, however, is pre-programmed[33] and can therefore be regarded as reflecting the deliberate and premeditated choices of programmers.[34] Someone has to decide in advance what a driverless vehicle will do in situations where the vehicle cannot avoid a crash altogether but can select what object or person it will collide with.[35] This opportunity for deliberation and premeditation implies greater moral responsibility[36] since a driverless vehicle's behavior in a crash reflects a deliberate choice among multiple options, rather than a reflexive reaction in the heat of the

[29] *See, e.g.*, Lin*, supra* note 19; Patrick Lin, *The Robot Car of Tomorrow May Just Be Programmed to Hit You*, WIRED (May 6, 2014), http://www.wired.com/2014/05/the-robot-car-of-tomorrow-might-just-be-programmed-to-hit-you/.

[30] Vehicular manslaughter charges typically require a level of negligence beyond ordinary negligence required for negligence torts. For example, some jurisdictions require "wanton or reckless disregard for human life" or negligence of a "gross and flagrant character, evincing reckless disregard of human life." 61A C.J.S. *Motor Vehicles* § 1671, at 329 (1970) (citing Burlas v. State, 971 A.2d 937, 943 (Md. Ct. Spec. App. 2009); Day v. State, 154 So. 2d 340, 342 (Fla. Dist. Ct. App. 1963)).

[31] FRAICHARD, *supra* note 9, at 3.

[32] Lin, *supra* note 29.

[33] By this, I of course do not mean that autonomous vehicle programmers could necessarily anticipate how they would act in any situation—but that the behavioral parameters of autonomous vehicles are human designs that aim to fulfill predetermined objectives and that reflect their designers' priorities.

[34] *See* Lin, *supra* note 19.

[35] *See id.*

[36] *See id.*

moment. These pre-planned choices can be examined, scrutinized, and regulated ahead of time.[37] As compared to human drivers, driverless car collision choices therefore bear a greater burden of moral justification, and greater opportunities for public scrutiny and accountability.

Google's current driverless vehicles use a combination of cameras, sonar, lasers, and radar to detect and identify obstacles around them.[38] These driverless vehicles can distinguish inanimate objects from pedestrians, bicycles, and other automobiles according to speed, geometry, and heat signature.[39] This is a necessary feature for a driverless vehicle's basic road navigation because identifying what an obstacle is enables the vehicle's computer to predict the obstacle's movement and behave accordingly: pedestrians move differently than bicyclists, cars, or stationary obstacles, so different precautionary measures are required to avoid collisions with each.[40] For example, bicyclists may be more likely to move erratically than tractor trailer trucks, which competent programming for driverless vehicles will need to take into account.[41]

With the ability to identify objects on and near the road, and the necessity of classifying them by type, programmers must choose how to allocate risk between the occupant of the driverless vehicle and other people on and near the road. While current driverless vehicles on the road today are reportedly programmed to simply minimize the overall probability of a collision without regard to allocating likely damage between itself and other objects, vehicles, and persons,[42] this would not be an optimal design in the long term. If a driverless vehicle

[37] In theory, the manufacturers of driverless cars could try to keep their "ethics programming" a secret. In practice, however, it is unlikely that this would be possible given the potential to reverse engineer the driverless vehicle computers. Regulators, insurance companies, litigants, and the public will likely demand to know how driverless vehicles are programmed.

[38] *See* Nitin Balodi, *Google Driverless Car—The Obstacle Detection Unit*, WHAT A FUTURE! (June 14, 2014), http://www.whatafuture.com/2014/06/14/google-driverless-car-the-obstacle-detection-unit/#sthash.NCdAM5F6.dpbs.

[39] *See* Nitin Balodi, *How Driverless Car Predicts Expected Movement of Objects on Road?*, WHAT A FUTURE! (Jan. 6, 2015), http://www.whatafuture.com/2015/01/06/google-driverless-car-predicting-movement-of-vehicles/#sthash.qQJG5eUs.dpbs. Google's current driverless vehicles do not have the ability to distinguish children from adults from elderly people or to determine how many people are in other vehicles, but these capacities are technically possible and could be implemented. *See* Jared Newman, *How to Make Driverless Cars Behave*, TIME (June 6, 2014), http://time.com/2837472/driverless-cars-ethics-morality/.

[40] *See* Balodi, *supra* note 38.

[41] *See id.*

[42] *See* Newman, *supra* note 39.

is on a collision path with an object, it would make sense for its programming to consider whether the risk to the vehicle is greater if it strikes the object or swerves. For example, if a vehicle on an icy road is on a path to collide with a traffic cone blown into its lane, it might be better to hit the traffic cone than to swerve to avoid it if swerving would present a risk of going off the road and hitting a tree. A general programming rule might be adopted where hitting some obstacle is preferred when swerving would risk hitting a more damaging object. If, however, the obstacle the vehicle is on a path to strike, or risks hitting if it swerves, is a pedestrian or another car, then it would make little sense for a vehicle to be programmed simply to select a trajectory that minimizes damage to itself. For an occupied driverless vehicle, failing to distinguish between pedestrians and traffic cones or trees and other vehicles would entail prioritizing the safety of its occupant over the safety of pedestrians and other motorists. In cases where an unoccupied vehicle is in an inevitable collision scenario, acting to minimize risk to itself without considering risk to others on or near the road would mean prioritizing its owner's property over other people's safety (or lives). That an unoccupied driverless vehicle should act in ways that avoid injuring pedestrians even at the expense of damaging itself, rather than preserving itself at the expense of running over pedestrians, should be an easy case to decide. Far more difficult cases, however, are presented when considering how much risk a driverless vehicle's programming should impose on its own occupants for the sake of protecting others' lives, bodies, and property.

Jason Millar introduced a thought experiment closely analogous to trolley problems[43] meant to illustrate the ethical dilemmas of driverless vehicle behavior during unavoidable collisions, which he termed the "Tunnel Problem":

> You are travelling along a single lane mountain road in an autonomous car that is fast approaching a narrow tunnel. Just before entering the tunnel a child attempts to run across the road but trips in the center of the lane, effectively blocking the entrance to the tunnel. The car has but two options: hit and kill the

[43] *See generally* Foot, *supra* note 28 (for the original account of the Trolley Problem); Judith Jarvis Thomson, *The Trolley Problem*, 94 YALE L. J. 1395, 1395 (1985) (for a classic analysis and exposition of the Trolley Problem).

> child, or swerve into the wall on either side of the tunnel, thus killing you. How should the car react?[44]

This scenario requires a choice between programming the car to kill its occupant or kill a child, both of whom are presumptively innocent.[45] This raises questions of how driverless vehicle programming should make these choices, what rationale should govern those decisions, and who should make these determinations. Unlike some trolley problems, the "tunnel problem" scenario does not provide any obvious utilitarian answer,[46] since the number of people killed would be the same under either choice. In classic trolley problems, many people's intuitions shift from wanting to minimize harm according to utilitarian considerations, to insisting on deontological constraints on how people can be used in cases where one person is instrumentally sacrificed to save others, such as in Thomson's "fat man" and "transplant" variants of the trolley problem.[47] The tunnel problem's choice, however, does not involve using either party instrumentally to save the other and lacks clear

[44] Jason Millar, *An Ethical Dilemma: When Robot Cars Must Kill, Who Should Pick the Victim?* ROBOHUB (June 11, 2014), http://robohub.org/an-ethical-dilemma-when-robot-cars-must-kill-who-should-pick-the-victim/.

[45] If perhaps the child appears "at fault" in this thought experiment, it is easy to conceive of modifications where they were not at fault and took every ordinary precaution. The MIT Media Lab launched a website, moralmachine.mit.edu, presenting a large series of variations on choices similar to the "Tunnel Problem" using a set up likely to occur more frequently in the real world than single-lane mountain tunnels: roads lined with barriers that make it impossible for a vehicle to drive off road, such that if a the path forward is blocked on one side by pedestrians and on the other side by an immovable obstacles, a vehicle that cannot break in time must either run over pedestrians or into an obstacle that will kill its passengers. The Moral Machine website also presents cases where a vehicle's path is blocked by multiple people at different places and the vehicle can avoid hitting some but not all of the people in the road. *See* Jacob Brogan, *Should a Self-Driving Car Kill Two Jaywalkers or One Law-Abiding Citizen?*, SLATE (Aug. 11, 2016), https://slate.com/technology/2016/08/moral-machine-from-mit-poses-self-driving-car-thought-experiments.html.

[46] Most seem to think that, in the original Trolley Problem, if a person is given the opportunity to divert a trolley on a track towards killing five people onto an alternative track where it would kill one person, they are either morally permitted or morally obliged to do so. *See* Thomson, *supra* note 43, at 1395.

[47] The "fat man" scenario is a case where, rather than diverting a trolley away from one track where it is headed towards five people, a man with sufficient weight to stop the trolley is pushed on the tracks to save five people. *See id.* at 1409. The "transplant" scenario is a case where a doctor has five patients in need of organ transplants who could be saved by killing one healthy patient for his organs. *See id.* at 1395–96.

act/omission distinctions because both choices amount to a programmed directive to kill one person to save the other.

The advent of driverless vehicles as a widespread form of transportation requires determining how they should behave in analogous scenarios. It is also necessary to consider cases involving different numbers of people—such as a car with two people in it headed towards one pedestrian or a car with one person in it headed towards two pedestrians. Cases involving asymmetric distribution of risk will also arise. For example, we could imagine a modified tunnel problem where, if the driverless car brakes immediately to mitigate the damage from the collision, the child has a 10% chance of surviving, but if the car swerves into the mountain, the motorist has a 20% chance of surviving.[48] Given how many millions of cars are on the road, it is reasonable to expect that every permutation of asymmetries in the number of people at risk and the severity of risk may arise in real world cases.

Current driverless vehicles are unable to distinguish numbers of occupants in other vehicles,[49] but such a feature would be necessary to implement in the future. When driverless vehicles gain widespread adoption, they are very likely to make a large portion of their trips without carrying any human occupants at all, such as when making deliveries, parking themselves remotely after dropping off passengers, or while driving to pick up passengers. Driverless cars that detect that they are not carrying any human occupants[50] should, I think uncontroversially, act to prioritize the safety of pedestrians and occupied vehicles rather than protecting themselves in inevitable collision scenarios.[51] To do otherwise would be to prioritize property over human life or bodily safety. The correlative of this is that driverless cars carrying human passengers ought to be able to recognize if other vehicles are unoccupied, and if so, behave in a way that maximizes the safety of their own human occupants at the expense of damage to unoccupied vehicles when necessary. The capability to prioritize occupied vehicles over unoccupied vehicles will likely imply equipping driverless vehicles with the capability to determine which of their seats are occupied and the numbers and

[48] In real-life scenarios, any predictions along these lines would be imprecise as to degree of confidence and margin of error.

[49] *See* Newman, *supra* note 39.

[50] Perhaps through heat sensors or weight sensors in seats.

[51] For example, an unoccupied driverless vehicle facing the Tunnel Problem should always swerve and spare the pedestrian at the expense of destroying itself.

positions of people riding in other vehicles.[52] As a result, it will be necessary to decide how, or if, driverless cars should factor in different numbers of people in each vehicle when allocating risk.[53]

Jared Newman identified another morally significant automobile collision scenario, where, rather than a driverless vehicle having to determine what object to strike, a vehicle might have to choose[54] between evading an oncoming vehicle altogether or enabling a collision that would mitigate the damage to the other vehicle.[55] Newman provides a hypothetical instance of such a case where a vehicle coming around a sharp bend on the side of a cliff loses its brake control, such that it will enter the path of a driverless vehicle.[56] In this scenario, if the driverless vehicle brakes hard, it can spare any risk of injury to its occupant, but the oncoming vehicle will careen over the cliff and likely kill its occupants.[57] If the driverless vehicle brakes softly, it could deliberately collide with the oncoming car such that the malfunctioning car will be spared from falling off the cliff face, but exposing the driverless vehicle's occupants to a wholly avoidable risk of injury.[58]

We might flesh out Newman's hypothetical by further specifying that in this thought experiment, the choice to make a deliberate,

[52] The ability to detect not only how many people are in other vehicles, but also which seats they are occupying, would likely enable safety enhancements in collision behavior since unavoidable collisions may be harm-minimizing by directing the vehicle's momentum against an unoccupied part of another vehicle's cabin rather than an occupied part.

[53] Perhaps the easiest way for driverless vehicles to account for the number of occupants in other driverless vehicles would be for each vehicle to transmit this information to a shared cloud network. An arrangement of this sort may however implicate privacy concerns, such as enabling the tracking of people's movements through driverless vehicles. This set of concerns was raised by Jack Balkin in correspondence on file with author. Manually driven vehicles, however, also compromise anonymity given that they are required to display easily photographed license plates and many states restrict the amount of window tint permitted. As such, technology enabling driverless vehicles to account for the number and position of occupants in other vehicles does not necessarily present new privacy concerns that are not already present given the ways the occupants of manually driven vehicles are observed and tracked using current technology.

[54] This phrasing is only meant to expediently describe paths available to a vehicle that its programmers might opt to favor or disfavor—of course driverless vehicles of the sort considered here do not have any agency of their own so they do not literally make choices. They may also be unable to carry out their programmers' preferences with perfect reliability given practical limitations.

[55] *See* Newman, *supra* note 39.

[56] *See id.*

[57] *See id.*

[58] *See id.*

damage-mitigating collision in this particular fact pattern will predictably result in fewer total fatalities over time when compared to programming that would allow the oncoming car to drive off the road, but the choice to execute damage-mitigating collisions would distribute the fatalities between the driverless vehicles' occupants and other motorists. In contrast, programming that leaves other motorists to their fate would allocate the probable fatalities entirely on the motorist with the faulty brakes.

Programming driverless vehicles to engage in damage-mitigating collisions might be termed the "altruistic crash case," one type of "occupant-indifferent behavior," that does not prioritize a person's safety merely because they are occupying the vehicle in question but instead exposes its own occupant to some risk in order to mitigate more dire risk to others. Programming driverless vehicles to allow the other motorist to drive off the cliff, or otherwise maneuvering in ways that prioritize the safety of their own occupants, could be classified as "occupant-favoring behavior."[59] Occupant-favoring behavior can be thought of as coming in different degrees according to how heavily the vehicle's behavior is biased towards the safety of the vehicle's occupants (or perhaps even more dramatically, preserving its own resale value). In the proceeding sections, this Article will address various occupant-favoring and occupant-indifferent models for programming driverless vehicles and consider how they impact different incentive structures and whether they produce fair results.

II. THE NECESSITY OF COLLISION BEHAVIOR REGULATION PRIOR TO THE WIDESPREAD ADOPTION OF DRIVERLESS VEHICLES

If government regulators do not implement uniform requirements for collision behavior programming, these ethically fraught choices will be left up to driverless car manufacturers and software designers. Depending on the technical particulars of consumer driverless cars, consumers might also be able to install or select their own collision behavior programming systems or modify their vehicles' original programming.

Enabling the choice of collision behavior programming in driverless vehicles to be made by their buyers and sellers would lead

[59] I have tried to use the terms "occupant favoring" and "occupant indifferent" rather than, for example, "occupant loyal," or "occupant biased," for the former and "egalitarian" or "utilitarian" for the latter so as to minimize the extent to which the terminology might manipulate a reader's intuition. I would, however, be happy to use more neutral terminology if any could be identified.

to a serious collective action problem. If consumers are able to choose ethics rules for their own vehicles, many are likely to choose those with highly occupant-favoring rules in order to understandably maximize their own personal safety.[60] However, people will also be unlikely to want to share the roads with driverless vehicles programmed to substantially prioritize themselves at the expense of others. Given that most people will also travel on foot, bicycle, in manually-driven cars, and in driverless cars subjected to the ethics programming of other driverless cars, if they are motivated to maximize their safety, it is reasonable to infer that most would choose occupant-indifferent programming as a rule of general applicability.

Yet, if left to the market, companies producing highly occupant-favoring driverless cars for consumers will have a decisive competitive advantage, since at the point of purchase, people motivated by safety maximization would do better to select an occupant-favoring car than an occupant-indifferent car. This will create a scenario where, if everyone wants to prioritize their own safety but can only act as consumers, their consumer preferences in selecting the safest cars for themselves (those that are occupant-favoring) will have the aggregate effect of making roads less safe for everyone, including themselves,[61] since occupant-indifferent programming could reduce total injuries and fatalities.[62] If left to the market, consumer demand would likely

[60] The popularity of SUVs is likely in part attributable to making consumers feel safer given their size and elevation, regardless of the obviously heightened risks that SUVs pose to others on the road. *See infra* note 83 and accompanying text. It might also be noted that external airbags to protect pedestrians were phased out of Volvos very shortly after they were first introduced. *See* Jeffrey N. Ross, *Volvo's Pedestrian Airbags May Already Be on Their Way Out*, AUTOBLOG (Dec. 1, 2013, 10:01 am), https://www.autoblog.com/2013/12/01/volvo-pedestrian-airbag-canceled. This would seem to suggest that there was little consumer demand for a safety device intended to protect people other than the vehicle's occupants.

[61] The extent to which a person will likely favor occupant-indifferent or occupant-favoring programming may depend on how often they walk or cycle, and how often they ride in driverless vehicles—an issue pointed out by Jack Balkin in correspondence on file with the author. However, even someone who rides in driverless vehicles as their exclusive mode of transportation would have an interest in others' driverless vehicles adopting an occupant-indifferent programming, even while having an incentive to select occupant-favoring programming for their own driverless vehicle if given the choice. Thus, even if a person who exclusively rides in driverless cars may want vehicle programming less protective of pedestrians than someone who always walks, they would still enjoy greater safety if other vehicles on the road adhered to occupant-indifferent programming while their own vehicle followed occupant-favoring programming. As such, even exclusive driverless car riders will face a collective action problem in favoring one rule of general applicability but preferring another rule if able to choose for themselves.

[62] *See generally supra* notes 57–61 and accompanying text.

result in collision behavior programming norms that are inconsistent with the public's preferences for collision behavior programming—where consumers would prefer one rule if choosing for their own vehicle, but a different rule if choosing for all vehicles, then programming norms driven by consumer choice rather than public governance will produce results undesirable even to the people making the occupant-favoring consumer choices.

Industry self-regulation presents an alternative model to public regulation. It is, however, an insufficient solution to this collective action problem because an incentive would remain to design collision behavior that favors the industry's customers as a class over people who are not their probable customers. Leaving driverless vehicle collision programming to industry regulation would, in effect, make public spaces shared with cars more attractive for driverless car users than for cyclists and pedestrians. This would, in some ways, mirror the evolution of traffic laws and social norms around road usage. As Peter Norton documents, city streets were historically shared by multiple modes of transportation,[63] but after significant social and political contestation, streets came to be understood as the privileged domain of automobiles.[64] Given the heightened scrutiny driverless vehicles will face as a new form of transportation, the manufacturers of driverless vehicles may have further incentives to optimize the safety of their occupants to instill confidence in their potential users. Even though the occupants of a driverless car programmed with occupant-indifferent collision behavior may be far safer than those using manually driven cars, the idea that a driverless vehicle would not maximize the safety of its own passengers may discourage already suspicious people from using them. Even if standards for collision programming are set by industry self-regulation rather than market competition, the industry is still likely to respond more the desires and fears of their customers than to the public health goal of minimizing total injury and loss of life. This is evidenced by the fact that manufacturers of manually driven automobile considered safety features for their drivers and passengers decades before considering measures to reduce pedestrian fatalities.[65]

[63] Such as pedestrians, private horse-drawn vehicles, streetcars, other city services, and children at play. *See* Peter D. Norton, *Street Rivals: Jaywalking and the Invention of the Motor Age Street*, 48 TECH. & CULTURE 331, 332 (2007).

[64] *See id.* at 341–47.

[65] *See Protecting Pedestrians Through Vehicle Design: Advancements Can Reduce Pedestrian Injury in Collisions*, EDMUNDS (May 4, 2009), http://www.edmunds.com/car-safety/protecting-pedestrians-through-vehicle-design.html. Google is researching external airbags to protect pedestrians for use in

One clear solution to this collective action problem is to adopt uniform government standards for collision programming. Government requirements for how driverless cars behave in inevitable collision states could resolve the collective action problem found in a consumer choice model by instituting programming predicted to reduce the total number of injuries and deaths.

Tort law and insurance might seem to be more obvious ways of regulating driverless vehicle collision programing than mandating parameters for that programming since human motorist behavior in collisions is largely regulated by insurance and tort law.[66] A substantial portion of the legal scholarship on driverless vehicle regulation has focused the implications of autonomous vehicles for tort and insurance law, of which there are certainly many interesting questions not taken up by this Article.[67] However, choices of insurance law will likely be insufficient to overcome the collective action problem considered here, and tort law is ill equipped to address it.

Three theories of tort liability govern automobile collisions today: negligence, no-fault liability, and strict liability.[68]

Negligence law makes little sense in the context of driverless vehicles. For an actor to be liable under a negligence theory, that actor's conduct must breach their duty to exercise reasonable care.[69] For a driver of a manually driven vehicle to be held liable for some damage under a negligence theory, the driver's conduct must have been unreasonable[70] (according to the standards of the jurisdiction) and a cause of the damage in question.[71] In a fully autonomous vehicle, where its operator has no contemporaneous control over steering, braking, or acceleration, the operator's conduct will not

its driverless cars, while Volvo also has a pedestrian airbag system. *See* Kukil Bora, *Google's Driverless Car Could Feature Air Bags on the Outside, Patent Filing Suggests*, INT'L BUS. TIMES (Mar. 25, 2015), http://www.ibtimes.com/googles-driverless-car-could-feature-air-bags-outside-patent-filing-suggests-1858458.

[66] JAMES M. ANDERSON ET AL., AUTONOMOUS VEHICLE TECHNOLOGY: A GUIDE FOR POLICYMAKERS 112 (2016) [hereinafter ANDERSON].

[67] *See, e.g.*, ANDERSON, *supra* note 66, at 111; Kyle Colonna, *Autonomous Cars and Tort Liability*, 4 CASE W. RES. J.L. TECH & INTERNET 81, 81–86 (2012); Daniel A. Crane et al., *A Survey of Legal Issues Arising from the Deployment of Autonomous and Connected Vehicles*, 23 MICH. TELECOM. & TECH. L. REV. 191, 256–95 (2017); David C. Vladeck, *Machines Without Principles: Liability Rules and Artificial Intelligence*, 89 WASH. L. REV. 117, 127–29 (2014).

[68] ANDERSON, *supra* note 66, at 112.

[69] RESTATEMENT (THIRD) OF TORTS: GEN. PRINCIPLES § 4 (AM. LAW INST., Discussion Draft 1999).

[70] *See id.*

[71] *See id.* § 3. The exact elements of negligence in automobile crashes and conditions where a driver is presumed negligent vary between jurisdictions.

normally cause any damage since the operator's conduct will not determine the vehicle's driving during a collision. If a vehicle operator's acts or omissions did not cause the harm in question, then considering whether their conduct was reasonable or negligent is inapplicable.

Strict liability standards in tort assign liability for harm without consideration of the defendant's negligence or intent.[72] There is no general rule of strict liability for physical harm in the sense that there is a general rule for negligence, but a set of doctrinal and statutory rules that assign strict liability in a variety of cases, each requiring different non-fault based elements.[73] One option for determining liability for collisions involving driverless vehicles would be to establish a doctrine that the manufacturer or user of the driverless vehicle is strictly liable for some or all of the resulting damage it causes.[74]

A strict liability rule for the damage caused by autonomous vehicles would create a legal regime where manufacturers and users of more dangerous vehicles (manually driven vehicles) can cause damage without liability unless they meet an additional fault-based element, but manufacturers and users of safer vehicles (future driverless vehicles) are held liable for the damage they cause regardless of fault. This would seem to be an unlikely doctrinal choice given that strict liability has traditionally been assigned to abnormally dangerous activities (like blasting[75]) or abnormally dangerous animals (like dogs known to bite),[76] where the defendant places others at some risk of physical harm even if they use reasonable care. Strict liability is also used for product defects, including defective product design[77] and defective manufacturing,[78] which would presumably apply to driverless vehicle manufacturers in the same way product defect liability applies to other manufacturers. In the case of driverless vehicles, the defendant's activities (whether a user or manufacturer is assigned the liability) would place others at less risk of physical harm than the risk posed by the activities of

[72] RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL & EMOTIONAL HARM ch. 4., scope note (AM. LAW INST. 2010).

[73] *Id.*

[74] For discussion of strict liability, see *id.* For an account of why strict liability for manufacturers of driverless cars might be an attractive tort option, see ANDERSON, *supra* note 66, at 116.

[75] RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL & EMOTIONAL HARM § 20 (AM. LAW INST. 2010).

[76] *Id.* § 23, cmt. e.

[77] RESTATEMENT (THIRD) OF TORTS: PRODUCTS LIAB. § 2 (AM. LAW INST. 1998).

[78] *Id.*

drivers and manufacturers of manually driven vehicles, so a strict liability rule would seem to impose the wrong set of incentives.

Strict liability for driverless vehicle collisions in a legal regime that uses negligence liability for manually driven vehicles would also seem to unfairly treat users of driverless cars more harshly for causing the same sorts of damage that manually driven cars cause, despite being less likely to cause it. Assigning strict liability to manufacturers of driverless vehicles but not manually driven vehicles might initially make sense because we would tend to assign blame for crashes to manual drivers before assigning it to automakers, whereas the manufacturer's choices alone could have contributed to driverless vehicle crashes. A driverless vehicle manufacturer's[79] choices would only be responsible for car crashes in a way parallel to the choices of traditional automakers though: certain design and programming decisions for driverless vehicles would predictably result in a certain distribution of harm, but the design choices of conventional automakers also predictably cause harms given their intended and expected uses. While collision programming choices, as argued in this Article, entail a deliberate choice of how to distribute harm, this is also the case for the design decision to make traditional cars that are intended to travel high speeds rather than low speeds, that are rigid rather than cushioned, that are heavy rather than light, or that have airbags on the inside for occupants but not on the exterior for pedestrians.[80]

The alternative answer for deciding whose insurance company should pay for physical harm caused by driverless vehicles would be to adopt a no-fault system, like the one used in twelve U.S. states for automobiles where damage under a certain threshold is assigned to the parties' insurance providers without regard for fault,[81] or a system like New Zealand's, where a private company or governmental agency covers everyone's personal injuries without consideration of fault.[82] This answer would address a different question than the one considered in this Article though: no-fault insurance is an answer to the question of who pays for physical harm, not the question of how unavoidable risks of physical harm should be distributed. Paying for medical bills or monetary compensation for physical suffering is not typically thought to fully address an injured person's losses, which are

[79] And the choices of their customers, because manufacturers would presumably price in the cost of added tort liability to the price of their vehicles.

[80] Or, for that matter, the choice to make cars rather than bicycles or airplanes.

[81] ANDERSON, *supra* note 66, at 113.

[82] *See What Your Levies Pay For*, ACC, https://www.acc.co.nz/about-us/how-levies-work/what-your-levies-pay/ (last visited Oct. 30, 2018).

in some ways incommensurate with money, and obviously, compensation for deaths cannot enjoyed by the decedents.

If government regulation of driverless vehicle inevitable collision programming is to take place, it must occur ahead of the widespread consumer adoption of fully autonomous vehicles. If regulation is deferred until driverless cars become a significant mode of transportation, then the industry, responsive to their existing consumers expectations, will likely be motivated to oppose such regulation. A certain portion of consumers will also want to be able to customize their driverless vehicle's collision programming or exercise some consumer choice over this aspect of their vehicles. If consumer choice in collision behavior becomes available, the availability of those choices will be seen as the socially expected baseline. Proposals to alter that baseline expectation by taking those choices away will be politically difficult, since this would likely be seen as undermining personal and consumer freedom—values that are especially culturally salient for automobile regulation. To use an example of the way market-driven vehicle design already prioritizes the safety of occupants over the safety of others, the size and height of SUVs likely makes their drivers feel safe and in control, even as these same features place pedestrians and drivers of smaller vehicles in greater physical jeopardy,[83] and SUV manufacturers have successfully prevented regulators from reigning them in.

An objection to uniform ethics rules requiring that driverless cars adopt occupant-indifferent collision behavior, raised by Patrick Lin, is that operators of driverless cars might reasonably expect that their vehicles "owe allegiance" to their owners and should therefore "value his or her life more than unknown pedestrians or drivers."[84] This makes sense if we presume that driverless cars will be regarded as if agents or fiduciaries of their operators. Perhaps this is a reasonable assumption, since people tend to believe that if they own something it should serve their needs above the needs of others. Whether people think of driverless cars this way or not, however, depends on what cultural expectations and attitudes come to surround them. If regulators can get out ahead of the widespread adoption of driverless cars to institute norm-shaping laws uniformly regulating their collision behavior, an absolute "loyalty" from one's car might not be expected (especially if driverless vehicles are most commonly used in

[83] *See* Malcolm Gladwell, *Big and Bad,* NEW YORKER, Jan. 12, 2004, at 29 (noting that, although they *feel* safer, SUVs are not actually safer for their occupants either).

[84] *See* Lin, *supra* note 19.

car hiring services rather than as privately-owned consumer goods). People expect some consumer goods to comport with their wishes in an utterly "loyal" way, like the ability to ride a bicycle anywhere and resell it, but do not expect all consumer goods to function this way, such as the expected inability to use a cable box to access all channels for free, or to resell that access. There is, however, often a belief that owners of consumer electronics have a "right to tinker," and they often do tinker with their computers and cellphones.[85] Nonetheless, under established law such as the Digital Millennium Copyright Act, consumer rights to modify software they've purchased are highly restricted,[86] so such restrictions on the software governing driverless car behavior would not be novel.

Automobiles are already a class of property that people are used to having highly regulated, from licensing and insurance requirements, to speed limits, to inspections, to special legal exposures while using them, to highly complex traffic laws.[87] Car usage is one of the only widespread activities that require users to frequently present themselves to a government office.[88] Collision behavior programming requirements might simply be regarded as another requirement of many for operating a car within the bounds of the law.[89]

[85] *See, e.g.*, John Black, *The Impossibility of Technology-Based DRM and a Modest Suggestion*, 3 J. TELECOMM. & HIGH TECH. L. 387, 396–97 (2005) (describing the view that consumers should have freedom to "tinker" and that digital rights management laws are inappropriate).

[86] *See* 17 U.S.C. § 1201 (2012).

[87] Examples include suspicion-less breathalyzer test requirements and the need to show a license and registration on demand. *See, e.g.*, 61A C.J.S. *Motor Vehicles* § 1575 (1970) (discussing statutes penalizing the refusal of a driver to submit to a breathalyzer test, noting that such statutes do not violate constitutional privacy protections); *id.* § 1696 (discussing statutes criminalizing the failure of a motorist to disclose personal information and identification).

[88] Specifically, a government office and set of bureaucratic interactions almost universally despised. *See, e.g.,* Ezra Klein, *Why Does the DMV Suck So Much?*, AM. PROSPECT (Mar. 24, 2009), http://prospect.org/article/why-does-dmv-suck-so-much.

[89] Where vehicles are already regulated, people do not seem to feel entitled to evade those regulations. For example, during the Volkswagen emissions scandal, where Volkswagen equipped its cars with software used to cheat on emissions tests, there was widespread outrage at Volkswagen and little outrage at emissions regulations. *See VW Emissions Scandal: Cheating and Outrage*, N.Y. TIMES (Sept. 21, 2015) http://www.nytimes.com/2015/09/25/opinion/vw-emissions-scandal-cheating-and-outrage.html. If people were to voluntarily install their own emissions cheating software, few would be very sympathetic to their claims of a "right to tinker." When modifying software or equipment harms non-corporate third parties, a "right to tinker" it is likely to find little support.

We might, however, anticipate another, more substantial objection to government-imposed requirements for driverless vehicle collision programming. Drivers of manually driven cars are given the legal and practical right to choose how to respond to inevitable collision states, including responses that prioritize their safety or even their property above others.[90] Why should the operators of driverless cars be allowed fewer freedoms and less control over their lives than those of manually driven cars, when they already sacrifice some of their freedom and control in a socially beneficial manner through opting for driverless vehicles?

Requiring occupant-indifferent collision rules could put the occupants of driverless vehicles at a disadvantage as compared to a manual vehicle driver in certain inevitable collision states. This would have the effect of requiring operators of driverless vehicles to trade off some of their own safety and freedom for an additionally heightened level of safety for those around them, when they are already reducing their relative risk to others as compared to manual vehicle drivers. Driverless vehicle operators might be thought to have already done their part to improve road safety as compared to drivers of manual vehicles—to extract yet a higher toll from driverless vehicles operators, at the expense of their personal safety, might seem to be itself an unjust allocation of risk.

Worse still, demanding occupant-indifferent collision programing could create the wrong incentive structure by discouraging people from using driverless cars out of safety concerns, even though the widespread adoption of driverless cars would likely improve overall safety, even if they were programmed to favor their occupants.[91] It would then seem to be worse, from a public safety perspective, to discourage people from using driverless cars as an alternative to

[90] Formally, this right is limited to civil and criminal negligence laws, but prioritizing one's own safety in an inevitable collision is unlikely to attract criminal or civil liability. Even if liability might attach, the practical ability to choose in a maximally self-preserving manner necessarily remains. It is not hard to imagine someone preferring to risk fines, a lawsuit, or even jail time to substantially reduce their risk of death or injury. A driverless vehicle operator who cannot customize his or her vehicle will be without this practical option.

[91] A completely rational safety-maximizing consumer would not be dissuaded from using driverless vehicles with occupant-indifferent programming on safety grounds since a driverless vehicle with occupant-indifferent programming could still be far safer for its own occupants than manually driven vehicles. However, consumer choices are frequently not fully rational, and a regulatory regime concerned with maximizing safety should be concerned with making people comfortable with driverless vehicles to encourage them to displace more dangerous, manually driven vehicles.

manually driven cars, even if the driverless cars they opted for were occupant-favoring.

Although these objections have some validity, they can be addressed in part by considering the particular nature of driverless vehicles. Drivers of manually driven vehicles have tremendous control of their vehicles by virtue of the vehicles' technical limitations: the vehicles cannot do the driving so human drivers are necessarily at practical liberty to make driving decisions. The operators of driverless vehicles are not in control of their vehicles in the same way because they are not doing the driving. As such, the degree of control that operators of driverless vehicles are afforded is not a technical necessity but a design choice. While manufacturers of manual vehicles have no technical alternative to allowing their drivers to decide how to steer, designers of driverless vehicles have options about what choices they give to their operators and can therefore be asked to explain those choices. The fact that there is no technical ability to require drivers of manual vehicles to consistently conform to uniform collision behavior rules does not suggest that we should not regulate driverless vehicle collision behavior given the technical ability to do so. Driving rules and norms developed over centuries given the political, technical, and financial limits of their time and are not necessarily the optimal baseline rules to begin with when considering a new legal regime for a new type of vehicle.[92]

The fact that the collision behavior of driverless vehicles must be dispassionately preprogrammed, rather than decided in the heat of the moment, also provides reason to subject them to greater regulation.[93] Given that driverless vehicle computers could model the physics of a collision and run the probabilities of injuries and fatalities in a way a human driver could not possibly do reliably,[94] the designers of driverless vehicles also govern behavior in inevitable collision scenarios with far greater knowledge than an ordinary human driver. Instinctual self-prioritizing by a human driver making a choice in less than a second with limited knowledge can be excused in the way that

[92] The fact that manually driven vehicles operate on a different set of legal baselines and can only be made so safe given their technical limits does not mean that this arrangement should be preserved. Instead, this arrangement may provide a reason to eventually phase out manually driven vehicles, at least in places where they expose people to the greatest dangers and offer the lowest recreational opportunities. This Article, however, aims to address the regulation of driverless vehicles before their widespread adoption, not to provide a general plan for addressing manually driven and driverless vehicles in the more distant future.

[93] Lin, *supra* note 19.

[94] I do not mean to suggest that this capability is currently available—but it seems like a reasonably likely computer and sensor capability given enough time.

the deliberate programming of a driverless vehicle, made ahead of time and with a potentially high degree of knowledge about outcomes, cannot be excused.

Nonetheless, as discussed in the next section, when a choice between manually driven cars and driverless cars remains available, there may be reasons not to simply maximize the safety returns from driverless vehicles when doing so might place an unfair burden of risk on their operators as compared to drivers of manual vehicles.

III. DIFFERENT WAYS TO PROGRAM A DRIVERLESS CAR FOR UNAVOIDABLE COLLISIONS

Having advanced a case for why it is necessary to impose uniform regulations on driverless vehicle collision behavior, it is necessary to identify and examine the ethical and public policy implications of adopting different regulatory regimes.

Jason Millar has argued that the only truly interesting question in the tunnel problem is not who the car should favor, but who should get to choose who the car favors, since any choice will be an arbitrary one without a correct answer.[95] Shifting the conversation from *what* should be decided to *who* should decide, however, does nothing to resolve the question of how risk should be distributed in tunnel problem type cases. This discursive move only defers this question to another decision maker rather than grappling with it directly. Any decision maker chosen to decide how driverless cars should behave in tunnel problems and equivalent scenarios will need to either provide a reasoned explanation for a choice or make an overtly arbitrary and unjustified decision. The following sections consider different systems for deciding who to prioritize in inevitable collision scenarios and what arguments can be mounted for and against them.

A. Strict Occupant Priority Systems

One straightforward option for collision behavior programming, implicitly suggested in Lin's notion that driverless vehicles might "owe allegiance"s to their operators[96] would be for driverless vehicles to prioritize the safety of their occupants without compromising it for the safety of others. In the tunnel problem, a car with strict occupant

[95] *See* Jason Millar, *Should Your Robot Driver Car Kill You to Save a Child's Life?*, CONVERSATION (Aug. 1, 2014), https://theconversation.com/should-your-robot-driver-kill-you-to-save-a-childs-life-29926.

[96] *See* Lin, *supra* note 19.

priority would always run over the pedestrian, braking to mitigate the damage to the extent possible but not putting its occupant at any added risk. In the altruistic crash case, the driverless vehicle operating with strict occupant priority would brake hard, protecting itself and letting the other vehicle drive off the cliff.

This rule of "strict occupant priority" could be justified by arguing that, since driverless vehicle operators are already exposing others on the road to far less physical risk than their manually driven counterparts, any additional safety dividends should accrue to the occupants of the driverless vehicles when their safety needs conflict with the needs of others. Such a rule of strict occupant priority would make riding in a driverless vehicle as safe as possible,[97] and in doing so, give people added reason to opt for driverless vehicles. To the extent that this reduces the amount of driving done manually, it would improve everyone's expected net safety. Strict occupant priority would also most likely mirror the default driving intentions of most human drivers, if they were able to drive perfectly. Driverless vehicles would then simply be augmenting human skill rather than performing a philosophically-motivated kind of public health intervention.

There are, however, significant problems with these lines of reasoning. Driverless cars can make everyone safer relative to manually driven vehicles, but the mere fact that someone has purchased or uses one should not necessarily provide a reason why their safety should be given greater priority than others when risk must be allocated between parties. Choosing to ride in a driverless vehicle will already provide enhanced safety for the occupants in the vast majority of cases—it is not clear then why it is necessary to allow their occupants the absolute safest ride, when doing so comes at the expense of greater risk for others. While it is sensible and usual to allow people to make purchases to enhance their own safety, it is not ordinarily thought acceptable to allow people to make purchases that enhance their own safety while relatively increasing the danger to others, even with respect to things that they own. A classic example of these norms is that it is generally lawful for people to build fences and install locks to protect their home and passively ward off trespassers, but it is illegal to install automated traps to further improve safety at the expense of danger to intruders.[98]

[97] This rule would apply except when in inevitable collision scenarios with other strict occupant-prioritizing driverless vehicles.

[98] *See* Graham Hughes, *Duties to Trespassers: A Comparative Survey and Revaluation*, 68 YALE L.J. 633, 649 (1959).

This line of reasoning might also be taken to suggest the conclusion that SUVs that enhance (or appear to enhance) their driver's safety at the expense of imposing greater dangers on other people should not be permitted. If it is not obvious why simply being the owner of a driverless car should grant a person greater safety dividends from an already net-risk reducing technology, it is even less clear why people should be able to expose others to greater risks while choosing a danger-increasing mode of transportation. Although SUV bans have been seriously considered by a number of municipalities,[99] such bans have not been widely adopted. Banning SUVs at this point would likely be politically impossible given that they are already widely used and available. This is, however, all the more reason to prohibit driverless vehicles from running heavily occupant-favoring programming before they come onto the market since, like SUVs, they would likely be desirable to consumers while carrying unjustifiable externalities.[100]

A more decisive objection to strict occupant priority systems is that, although driverless vehicles can expose others to much lower risks than manually driven vehicles even if programmed to absolutely favor their occupants, such that they could be seen as already exceeding the expected safety baseline, they still expose other people to greater jeopardy than cyclists and pedestrians. Any automobile on the road puts cyclists and pedestrians at an asymmetric risk of death and serious bodily injury since cyclists and pedestrians do not expose automobile users to equivalent jeopardy. Cycling and walking are also activities that are better for both the local environment and climate than driving,[101] contribute positively to public health, and incur far lower road maintenance costs to be shouldered by the public. It would

[99] *See, e.g.*, Nick Kurczewski, *Paris Considers a Ban on SUVs*, N.Y. TIMES (Jan. 13, 2011), http://wheels.blogs.nytimes.com/2011/01/13/paris-considers-a-ban-on-s-u-v-s/; Andy Bowers, *California's SUV Ban*, SLATE (Aug. 4, 2004), https://slate.com/news-and-politics/2004/08/california-s-secret-suv-ban.html.

[100] That SUVs are currently politically impossible to prohibit does not mean that they could not be phased out if appealing alternatives were made available or if the norms of acceptable risk for travel changed with the adoption of driverless vehicles. Likewise, the fact that a product might be desirable to consumers does not make it impossible to ban. Automobile regulations have the effect of banning the production of many vehicle designs that consumers might want, such as cars that can be made more cheaply by failing to meet safety and emissions standards.

[101] Even electric vehicles represent a far greater carbon load than walking and cycling, given their manufacturing and energy production needs. *See, e.g.*, Nuri Cihat Onat et al., *Conventional, Hybrid, Plug-in Hybrid or Electric Vehicles? State-Based Comparative Carbon and Energy Footprint Analysis in the United States*, 150 APPLIED ENERGY 36, 40–47 (2015) (describing the results of a comparative analysis of electric vehicles, plug-in hybrid vehicles, and hybrid electric vehicles.).

then make sense to think that if we want to incentivize people to choose driverless cars over manually driven cars, we will still want to incentivize people to choose walking or cycling over automobile transportation when possible.[102] Given that the occupants of driverless vehicles will already pose a greater danger to cyclists and pedestrians than vice versa, we should not permit their collision behavior programming to allocate risk further in favor of driverless car operators than inherently necessary. To do so would, in effect, give driverless car operators an unnecessary double safety advantage over cyclists and pedestrians during collisions—the inherent advantages of being incased in a protected, heavy vehicle, plus the added advantage of programming meant to prioritize the vehicle's occupants. When considered with the fact that the potentially lethal element in a vehicle-pedestrian collision is typically the vehicle and not the pedestrian, this would seem to be an imprudent and unjust allocation of risk by enhancing the already safer party at the expense of the more vulnerable party's safety when the former was the one responsible for creating the danger.[103]

B. Biased Priority Systems

A hypothetical collision behavior rule that can be used to illustrate the perils of allowing user customization might be described as "biased priority." A driver of a manual vehicle who would choose to run over a stranger in the tunnel problem might choose instead to sacrifice themselves if the pedestrian in the tunnel problem was their child or spouse.

We could imagine, in a legal regime where people were able to customize their vehicle's collision behavior programming, some people might upload a list of photographs and license plates of their family members that their vehicles would give special priority to upon recognizing via facial recognition or license-plate reading software, should they be involved in an inevitable collision state. Perhaps bigoted driverless vehicle operators, if free to tinker with their

[102] This will serve as the basis for the "incentive-weighted system" described later in this Article.

[103] At least in the intuitive sense that the kinetic energy of the potentially lethal collision is overwhelmingly that of the car and not the pedestrian—of course the presence of the pedestrian is also a but-for cause of a pedestrian-automobile accident, but as between a heavy fast moving object and a slow moving person, we normally would not think the proximate cause of the collision was the person merely being in the way.

vehicle's collision behavior, could specify people of certain demographics to be favored over others.[104]

These choices would probably be rare even in a regulatory regime that permitted them, and they are always a hypothetical possibility for drivers of manually driven vehicles.[105] From the standpoint of a democratic state committed to equal citizenship and the elimination of discriminatory treatment, however, there is good reason to prohibit such choices. While people are not equally physically safe, the government should seek to suppress deliberate efforts to expose some people to greater physical risk than others where possible. Preventing the use of biased priority collision behavior programming then provides another reason for prohibiting user-modifiable collision behavior programming.

C. Utilitarian Systems

Perhaps the most obvious system of collision behavior programming would be to require driverless vehicles in inevitable collision states to take actions that would cause the fewest fatalities and injuries on average. Noah Goodall described a system rewarding "behaviors that minimized global damage" which he termed "rational ethics" as his preferred rule for driverless vehicles in the early stages of their deployment.[106] This system might more precisely be termed a utilitarian approach, since other ethical systems could be described as "rational," but this system specifically seeks to quantitatively minimize the damage caused by collisions.

Under a utilitarian system, a driverless car faced with a tunnel problem scenario would drive over a single pedestrian if two people were in the car and swerving would, on average, result in more than one death. A car governed by a utilitarian system, however, would swerve if a single person occupied it and the pedestrian would be less likely to survive being hit by the car than the occupant would be to survive running into the tunnel wall, since this behavior would result

[104] In fact, participants in the "moral machine" experiment do show demographic preferences in tunnel problem like scenarios. *See* Edmond Awad et al., *The Moral Machine Experiment* 563 NATURE, 59–64 (2018).

[105] Of course, true inevitable collision states that entail choosing how to distribute risk between different people are likely to be extremely rare in proportion to the total amount of both miles driven and other types of accidents—but given very large numbers of vehicles over very long periods of time we can expect that they will sometimes occur.

[106] Noah J. Goodall, *Ethical Decision Making During Automated Vehicle Crashes*, 2424 TRANSP. RES. REC. 58, 63 (2014).

in less loss of life over time. A driverless car operating under a utilitarian system would likewise put itself in harm's way during an altruistic crash case if the expected total injury and loss of life with a damage mitigating collision between all participants would be less than the expected loss of life by avoiding the crash.

The obvious advantage of the utilitarian system is that, presuming no effect on the ratio of driverless vehicles to manual vehicles on the road,[107] it should result in the fewest automobile fatalities and injuries of any collision behavior system over time. If maximizing safety is the goal, a numerical utilitarian system is the clearest way to do it.

A utilitarian distribution of harms and benefits, however, is almost never the rule in our society. Why should we make it the rule for driverless vehicles? We do not, for example, require that distribution in healthcare, taxation, or property follows a utilitarian regime.[108] Perhaps we should, but why would we start with driverless vehicles? To do so would have the effect of risking some driverless vehicle operators' lives in situations where more lives can be preserved by putting their's in greater jeopardy. In a society that lacks even a duty to rescue when no risk is involved,[109] this would seem to be an unusual burden.[110]

One difficulty would arise in how a utilitarian system would handle cases where it has to allocate damage to one person or another but an insufficient amount of information is available to determine which choice would result in greater damage or probability of death. Suppose that in a tunnel problem case, the driverless vehicle cannot

[107] The benefits of this system might be undone if it sufficiently discourages adoption of driverless cars by people who would otherwise use manually driven cars.

[108] Different versions of utilitarianism—such as preference fulfillment maximizing utilitarianism as compared to welfare maximizing utilitarianism—might each call for very different distributive schemes but few areas of public policy would satisfy any of them.

[109] *See generally* Marin Roger Scordato, *Understanding the Absence of a Duty to Reasonably Rescue in American Tort Law*, 82 TUL. L. REV. 1447, 1452 (2008).

[110] Jack Balkin has suggested (in correspondence with the author) that this criticism of utilitarian programming could be mobilized against the other collision behavior programming rules described in this Article. That utilitarianism is rarely required, however, does not mean that self-serving behavior is always permissible. If someone physically harms another, it is not generally an excuse that doing so mitigated the danger posed to oneself. The absence of a duty to save another who is in danger does not apply to someone who created the danger themselves. Thus, while it would be unusual to require strictly utilitarian behavior it is not so unusual to require that people internalize and mitigate the risks they create. This observation supports a preference for what I will later describe as "incentive-weighted utilitarian" programming.

tell whether its own occupant or the pedestrian is more likely to survive a choice adverse to their safety. Or, perhaps the driverless vehicle could determine that the pedestrian has a near certain chance of death if the car merely brakes, and the car occupant has a near certain chance of death if the car swerves. How does it choose which to favor between the two on utilitarian grounds? A strictly numeric utilitarian standpoint provides no guidance on which is preferable. It could have an algorithm that makes an arbitrary choice, but this would lead to extremely unsatisfying explanations for why a car killed or maimed a particular person.

One partial solution to this would be a different utilitarian rule: rather than seeking simply to minimize total damage, instead seek to maximize total expected quality adjusted life years (QALY)[111] of the parties to an inevitable collision state. Such a rule would allow for a driverless car to run over a pedestrian in the tunnel problem if the pedestrian was significantly older than the car's occupant, but would require that, if it was carrying an elderly person that the car should swerve into the tunnel wall to save a child pedestrian.[112] Or, perhaps a car programmed in such a manner would even kill two elderly occupants to save a single child if the combined remaining life expectancy of the two occupants was less than that of the child as estimated by the driverless vehicle. Given potential advances in facial recognition software, this would likely be a technical possibility, at least in some cases.

The advantages of such a system over a numerical utilitarian system would include maximizing the total QALYs remaining after inevitable collisions given the limits of the vehicles' technical abilities to estimate and maneuver. This goal would be consistent with some, although not all, widely accepted approaches to public health. A QALY utilitarian system, however, would entail age (and possibly disability) discrimination, and as such, should be rejected for the same reasons the biased priority system should be rejected—it violates principles of equal treatment between similarly situated persons.

A deeper issue with implementing a utilitarian system is that, because such a system could only control the behavior of driverless vehicles rather than all road users, it could have lopsided risk distribution effects. Given that manually driven vehicles would not be

[111] *See, e.g.*, Franco Sassi, *Calculating QALYs, Comparing QALY and DALY Calculations*, 21 HEALTH POL'Y PLAN. 402, 402–03 (2006) (describing the method by which QALY is calculated).

[112] This could be a rationale behind the "moral machine" participants apparent preference against elderly people and in favor of babies and children. *See* Awad et al., *supra note* 104.

governed by utilitarian motives during collisions, requiring a utilitarian system for driverless vehicles during inevitable collision states could create a situation where occupants of driverless vehicles are at a greater proportional risk than manually driven vehicles in certain scenarios. This is because certain scenarios will dictate that a utilitarian driverless car sacrifice itself but a manually driven car would presumably not act likewise if presented with identical circumstances. It would also be extraordinarily unlikely for any manual driver to ever have the wherewithal and skill to execute a damage mitigating crash, but driverless vehicles with near perfect driving precision and modeling of collision physics might be able to do so in some rare cases. When some vehicles follow utilitarian rules and others follow self-prioritizing rules, to the extent that it is possible to allocate risk, risk will be unevenly distributed in favor of those with self-prioritizing behavior when adjusting for the advantages of driverless vehicles' accident avoidance abilities.

It might also be the case that a utilitarian system further amplifies some of the moral hazards for how other drivers might conduct themselves with driverless vehicles. Patrick Lin has pointed out that if people understand the crash avoidance system of driverless vehicles they could exploit it—for example, by aggressively cutting off a driverless vehicle knowing it will brake or swerve safely with a level of reliability that could not be expected from a human driver.[113] If utilitarian programming was generally known, then drivers of manually driven vehicles could maneuver aggressively around driverless vehicles with the knowledge that, at least if they are carrying more passengers, the driverless vehicle will sacrifice its own safety for theirs. A utilitarian system may, as a result, place driverless vehicle users in a position where they are vulnerable to exploitation and where some manual drivers may be tempted to engage in more aggressive driving.

Finally, from the standpoint of the pedestrian or cyclist, a utilitarian collision rule could also seem unjust. There could be cases, such as the tunnel problem involving two driverless car occupants and one pedestrian, where a driverless car with utilitarian programming would choose to kill a pedestrian to spare multiple car occupants. Yet in such a scenario, it would have been the transportation choice of the driverless car occupants that created the lethal danger in the first place—the kinetic energy of a fast-moving vehicle presents inherent dangers to its occupants and those around it, whereas the danger the pedestrian poses to the driverless car would only be a product of

[113] *See* Lin*, supra* note 19.

redirecting that kinetic energy away from them. Why should the driverless car occupants in such a scenario be able to both create the risks, and then benefit from having those risks allocated onto someone else, rather than having to internalize the danger created by using an automobile? The utilitarian system may then seem to fail at least some standards of fairness. However, as described in the next section, this argument requires more work to overcome the Coasean critique of causation.

D. Coasian Considerations

When considering who is responsible for the dangers in the tunnel problem, the motorist or the pedestrian, it is necessary to address the difficulties identified by Ronald Coase concerning how to attribute harm. Coase argued that the traditional approach of thinking that when A harms B, a liability rule should be adopted to restrain A obscures the reality of the choice of whether A or B should be privileged to harm the other.[114] Coase writes:

> The question is commonly thought of as one in which A inflicts harm on B and what has to be decided is: how should we restrain A? But this is wrong. We are dealing with a problem of a reciprocal nature. To avoid the harm to B would inflict harm on A. The real question that has to be decided is: should A be allowed to harm B or should B be allowed to harm A? The problem is to avoid the more serious harm. I instanced in my previous essay the case of a confectioner the noise and vibrations from whose machinery disturbed a doctor in his work. To avoid harming the doctor would inflict harm on the confectioner. The problem posed by this case was essentially whether it was worth while, as a result of restricting the methods of production which could be used by the confectioner, to secure more doctoring at the cost of a reduced supply of confectionery products.[115]

If a driverless car strikes and kills a pedestrian, we could describe the situation as one where the car *caused* the pedestrian's death by traversing the space where the pedestrian stood, or the pedestrian

[114] Ronald H. Coase, *The Problem of Social Cost*, 3 J.L. & ECON. 1, 2 (1960).

[115] *Id.*

caused her own death by standing there. Both the driverless car's actions and the pedestrian's actions are but-for causes of the fatality. While the pedestrian in such a scenario would have survived if the driverless car operator chose to walk instead, the pedestrian would likewise have survived if she chose to drive rather than walk. To require a driverless car on a collision course with a pedestrian to swerve into danger[116] would be to allow the pedestrian to cause harm to the driverless car occupant, since the pedestrian's presence would cause the car to swerve. The driverless car collision scenarios therefore closely map onto Coase's argument that tort liability rules represent choices between reciprocal harms.[117]

If we cannot choose which harms to avoid through appeal to causation or responsibility, and unmodified utilitarian rules are problematic, risks in collisions must be allocated according to other considerations.[118] Even if both parties in a collision bear but-for responsibility, one may bear greater moral responsibility by choosing a less desirable course of action. The following two sections offer alternative ways of modifying utilitarian collision programming to take into account different sorts of morally or socially relevant responsibility. In the first, which I call "fault-weighted utilitarianism," legal fault for an accident might serve as a type of tiebreaker for allocating risk. In the second, which I call "incentive-weighted utilitarianism," transportation choices which in the aggregate minimize risk and cost are prioritized relative to those which, on average, carry greater risks and costs. Neither system relies on attributing privileged causation, but instead makes a judgment as to which harms should be minimized.

E. Fault-Weighted Utilitarian

One modification of a utilitarian system that would reduce some of the moral hazards and compensate for the fact that it could be applied only to driverless vehicles and not manual ones would be to factor in legal fault when it comes to risk allocating decisions. Where all participants are faultless, a driverless vehicle acting according to

[116] This is the only really relevant case, since a driverless car on a course to strike a pedestrian that is able to successfully brake or swerve without endangering its occupant should obviously do so.

[117] *See generally* Coase, *supra* note 114, at 2.

[118] Even if Coase's argument on causation is accepted, this is clearly not a situation where the Coase theorem, even if it were not otherwise problematic, could apply, since there is no possibility of people bargaining over driverless vehicles' collision programming.

a fault-weighted utilitarian system would behave in a utilitarian manner, but when one party was responsible for causing a collision that would have been avoidable absent their fault, driverless vehicles operating under such a system would proportionately discount their safety in relation to the safety of the faultless party.[119] While this might be characterized as a sort of "technological due process,"[120] insisting that driverless vehicles sacrifice their occupants to protect people whose blameworthy actions caused the danger in question seems unfair if it is technically possible to do otherwise.

When costs must be allocated between two parties, but one party's blameworthy actions incurred those costs, it would often seem to be unjust to allocate those costs evenly. This is not, however, to say that costs should be foisted entirely onto the party at fault, only that they should be weighted. It would still be unreasonable for a driverless vehicle to protect its occupant from mere financial harm by taking actions likely to cause the death of the person at fault in an accident. Such a weighting would clearly be disproportionate in relation to any fault. Instead, in a scenario where a driverless vehicle must choose whom to expose to equivalent risk, such as a tunnel problem where braking and swerving both incur a high risk of death but assign it to different people, favoring the person who was faultless would at least plausibly create a better incentive structure than ignoring fault.

In a fault-weighed system, a driverless vehicle could hit a pedestrian in the tunnel problem if the pedestrian was acting in a negligent manner in violation of local traffic laws, or perhaps discount the priority afforded to the negligent pedestrian in relation to the negligence rather than strictly in proportion to the pedestrian's survival odds. If, however, the pedestrian was faultless, the vehicle would behave in a utilitarian manner. Likewise, in the altruistic crash case, if the other vehicle in peril was at risk due to the driver's own negligence, the driverless vehicle operating under a fault-weighted utilitarian system would not seek to rescue the driver from her own negligence at the expense of its occupant's safety. If, however, the out of control vehicle in the altruistic crash case was in danger due to no fault of her own, the driverless vehicle would act according to utilitarian calculations.

[119] It would probably be very technically challenging for a driverless vehicle equipped with current sensors to determine fault in many collision scenarios, so this is likely a mostly theoretical system rather than a practical one.

[120] A characterization offered by Jack Balkin in expressing skepticism over this proposition.

A fault-weighted utilitarian system would have the advantages of a utilitarian system without some of its disadvantages. It would provide the right incentives to manual car drivers to avoid accidents rather than providing the moral hazard of an ethics system that might tempt other motorists to drive more aggressively given the belief that driverless cars would compensate for their recklessness.[121] A strictly utilitarian system that makes no judgment of fault would have the terrifying vulnerability that two malicious people could deliberately kill a single driverless car occupant by jumping into its path, in a situation where the environment creates a tunnel problem scenario, since they would know that the utilitarian driverless car would swerve to its occupant's doom rather than run over two people.[122] While this may seem farfetched, any programming that gives other people a reliable and probably difficult-to-prosecute way of murdering its users could stand to have that vulnerability closed. A fault-weighted system could fix that problem.

A fault-weighted utilitarian system, however, would still have other problems found in the strictly utilitarian systems. Such a system would still have lopsided distribution effects given that manually driven vehicles would be operating without utilitarian concerns. Such a system would also still expose cyclists and pedestrians to danger out of proportion to the danger cyclists and pedestrians cause to others.

F. Incentive-Weighted Principles

I would suggest remedying the problems posed by pedestrians, cyclists, and drivers of manual vehicles in the previously discussed collision behavior systems through what I will call an incentive-weighted system. The incentive-weighted system asks the question without regard to fault: which party's transportation selection and behavior, if adopted more widely, would reduce the risks and costs of travel? When such a party can be identified, and someone must be placed at heightened risk through the driverless vehicle's actions or inactions, the driverless vehicle should prioritize the party whose transportation choices and behavior, if more widely adopted, would most reduce the risks and costs of travel. This collision rule would

[121] Driverless vehicles might all have the effect of compensating for others' recklessness and therefore encourage certain aggressive driving in cases where they can avoid collisions altogether.

[122] If the choice is between permitting such a scenario and engaging in a kind of "technological due process," the later seems less objectionable.

allow people to enjoy the benefits of their risk-reducing choices, when possible, and prevent people from externalizing the heightened risks incurred by their transport choices when those risks cannot be eliminated.

In this system, if a driverless vehicle was in an inevitable collision state with a manually driven vehicle due to the manual driver making an error that a driverless vehicle would not, the driverless vehicle should be able to prioritize its own occupant if it must choose who to risk. This rule would deny manual drivers the opportunity to free ride on driverless vehicles' collision programming, such that the dangers they pose to driverless cars that driverless cars cannot completely eliminate are distributed to the manual drivers themselves before they are distributed to driverless car occupants if this distribution is within the control of the driverless car's programming. This would have the effect of requiring manual drivers to internalize the excess risk they pose to driverless vehicles when that risk cannot be eliminated but could be shifted partially back onto them. If such a rule of priority is not implemented, then the design of driverless cars would be such that driverless car occupants will have to disproportionately absorb the risk that manual drivers create, rather than deflecting it back to the riskier vehicles.

If a driverless car was in an inevitable collision state with another driverless car and could choose in part how to allocate risk between them, the two cars should adopt a utilitarian formula. This is because the transportation choice of each occupant exposed the other to a symmetrical risk and, being in a driverless vehicle, each bears no personal responsibility for the circumstances of the collision.

A driverless car following an incentive-weighted system would have to behave differently when accounting for pedestrian and cyclist collisions, however. In places where walking and cycling are plausible options, they are the options that, when adopted more widely, reduce costs and risks as compared to using automobiles of any sort. A driverless vehicle adhering to an incentive-weighted system in an inevitable collision state with a pedestrian must prioritize the pedestrian because the choice to drive rather than walk, in the aggregate, increases travel-related fatalities. Had both chosen to walk, the overall danger from travel would be diminished to a greater degree than if both chose to ride in a driverless vehicle. To allocate risk otherwise would be to make the pedestrian pay for the driverless vehicle operator's more dangerous transportation choice. Similarly, a driverless vehicle failing to prioritize its own occupant in collisions involving erroneously-driven manual cars entail forcing the driverless vehicle's occupant to sustain costs the manual driver created.

An exception might be found in situations where pedestrians are on highways that are not intended to be navigable on foot, such as multi-lane, elevated intercity highways. While walking should be incentivized in urban and suburban settings, it is impractical to encourage people to walk over long distances. Motorized transportation of one sort or another is the only way to travel long distances without incurring socially unacceptable time costs. High-speed intercity highways are almost certainly more dangerous[123] with some pedestrians attempting to travel on them than with no pedestrians. Unlike city streets in densely populated areas, it is not practically necessary for pedestrians to use intercity highways,[124] and there would be dramatic social costs if automobiles could not use highways. As a result, walking on highways, unlike walking in cities, does not represent a safe mode of transportation that ought to be incentivized. An incentive-weighted system would then not require that a driverless car give pedestrians priority in risk allocation in highways.[125]

Just as there is an unfairness in forcing a typical pedestrian to bear the physical cost of someone's choice to use an automobile rather than walking, it would be unfair to sacrifice a driverless vehicle operator to compensate for the unreasonableness of a walking in a highway designated for use by automobiles only. An example that illustrates this point is that, if a suicidal pedestrian leaps in front of a driverless car and the driverless car can choose only between hitting the suicidal pedestrian and swerving off the road precariously near a cliff, the driverless car operator should not have to risk death in order to save someone who deliberately placed themselves in danger.

An incentive-weighted system would, therefore, answer the tunnel problem, according to whether the pedestrian created an unnecessarily heighted hazard by walking in a place inherently

[123] In the sense that there will be more fatalities on a highway if some amount number of pedestrians try to travel on it than if it is exclusively used for cars. An intercity highway would be safer still, presumably, if only pedestrians traveled on it, but this would defeat the purpose of having designated high-speed intercity highways—pedestrians can walk between cities without using highways should they choose to do so.

[124] The entire reason why intercity highways are desirable for automobile travel is that they permit high speeds without stopping. Since pedestrians cannot achieve high speeds anyways, intercity highways do not lend themselves to pedestrian travel over surface roads.

[125] Of course, if a driverless vehicle could avoid a fatal collision with a pedestrian by exposing its occupant to only financial or non-fatal risks, it should still do so. Priority weighting should only take place when a driverless vehicle must choose between likely causing roughly equivalent harm between people.

unsafe for pedestrians, or the driverless car operator chose the more dangerous form of transportation by being driven when she could have walked. The incentive-weighted system would resolve the altruistic crash case such that a driverless car should not put its occupant in harm's way to mitigate the danger to a manually driven vehicle, since the manual driver's choices, if aggregated, heightened the dangers on the road. As such, the driverless car operator should be permitted to benefit from their contribution to reduced transport danger. However, if a driverless vehicle finds itself in an altruistic crash case with another driverless vehicle, an incentive-weighted approach would require it to make a damage mitigating collision. This is because in such a case, both occupants' choices, in the aggregate, would have had the same predicted effect on road safety and the position of each in the collision scenario would be a result of brute luck, not more or less safety promoting choices.

The incentive-weighted system rewards transportation choices that reduce danger in the aggregate by prescribing collision programming that prioritizes their safety relative to those adopting less aggregate safety-promoting choices. It also avoids a programming system where driverless vehicles absorb excess risk from those making more dangerous transportation choices and engaging in riskier behavior on the road. This encourages safety-maximizing choices without the potential for the unfair uneven distribution of risk as between individuals found in a purely utilitarian system.[126]

In this way, the incentive-weighted system could be justified in both consequentialist grounds and Kantian grounds. The incentive-weighted rule would provide the right incentive structure to encourage people to adopt safer driverless vehicles rather than more hazardous manually driven cars by eliminating the possibility that

[126] The case presented in this Article is that contribution to risk should count for something, but it does not resolve the question of *how much* it should count when there are other factors that would strongly favor an alternative decision, such as differences in number of people at risk or differences in the amount of risk different parties would be exposed to. The incentive-weighted system described here would answer the question of how to "break a tie" between the equal interests of different individuals at risk. What it does not accomplish, however, is how or if transportation choice should be factored into cases where an unequal number of people are placed at risk. For example, when considering how to allocate risk in an inevitable collision state between a driverless car with one occupant and a manually driven bus with thirty occupants, that opting for a driverless car improves safety over opting for a manual car does not credibly outweigh a large disparity in numbers—or discount the social and environmental benefits of adopting mass transportation. Additional arguments would be required to propose relative weights for each relevant factor so that a "break even" point could be identified.

choosing a driverless vehicle could expose them to risks from manually driven vehicles or malicious actors. The incentive-weighted scheme would also further insulate cycling and walking from the hazards of automobile use, thereby discouraging those preferable modes of transportation to a lesser degree than a strictly utilitarian system. From a deontological perspective informed by Kantian considerations, the incentive-weighted system would respond to the duties to other's safety implicated in selecting more hazardous modes of transportation, while avoiding penalizing people for making choices that others would have reason to want to see universally adopted.

CONCLUSION

However rare inevitable collision states may be for driverless vehicles, with enough driverless vehicle on the road they are certain to occur in significant numbers over time. At least some of these inevitable collision states will place driverless vehicles' programmers in the position of determining how to allocate risk and expected damage between different parties. If left to consumer choice or industry regulation, driverless car collision behavior is likely to be inconsistent with the overall public good and aggregated public preferences. As such, government regulation is necessary to overcome what would otherwise be a collective action problem.

It is of little help to merely suggest that the government should regulate driverless car collision behavior without explaining how it should be regulated. The two most widely discussed collision behavior systems, which might be described as occupant-favoring and utilitarian systems, create moral and prudential problems when dealing with manually driven vehicles, pedestrians and cyclists. An alternative system, which I term an "incentive-weighted system" fills in these gaps and addresses many of these problems. The incentive-weighted system has the effect of encouraging all parties to minimize the dangers that they create by their choice of transportation. This corrects for both the lopsidedness of having driverless vehicles with preprogrammed collision behavior sharing the road with presumptively self-protective drivers of manual vehicles as well as the fact that driverless vehicles would still pose asymmetric dangers to pedestrians and cyclists.

The ethical dimensions of driverless car collision behavior will have to be programmed by the time fully autonomous driverless vehicles are in widespread consumer use. It is important to determine a rational and fair collision behavior system now, before these problems arise, to make sure that societal interests, public health, and interpersonal fairness are reflected in the rules for this new form of transportation.