


Crashing into the Unknown: An Examination of Crash-Optimization Algorithms Through the Two Lanes of Ethics and Law

Jeffrey Gurney

Related papers

[Download a PDF Pack](#) of the best related papers 



[Sue My Car Not Me: Product's Liability and Accidents Involving Autonomous Vehicles](#)
Jeffrey Gurney

[Three Types of Structural Discrimination Introduced by Autonomous Vehicles](#)
Hin-Yan Liu

[The Coming Collision Between Autonomous Vehicles and the Liability System](#)
Gary Marchant

CRASHING INTO THE UNKNOWN: AN EXAMINATION OF CRASH-OPTIMIZATION ALGORITHMS THROUGH THE TWO LANES OF ETHICS AND LAW

*Jeffrey K. Gurney**

I. INTRODUCTION

One, a[n] [autonomous vehicle] may not injure a human being, or, through inaction, allow a human being to come to harm. . . .

Two, . . . a[n] [autonomous vehicle] must obey the orders given it by human beings except where such orders would conflict with the First Law. . . .

Three, a[n] [autonomous vehicle] must protect its own existence as long as such protection does not conflict with the First or Second Laws.¹

Isaac Asimov first penned his three rules of robotics in a short story titled “Runaround.”² These rules formed the governing principles of the robots in his stories, and also served as a plot device to show the shortcoming of programming robots to follow these three simple

* Law Clerk, The Honorable Timothy M. Cain, U.S. District Judge for the District of South Carolina. J.D., 2014, University of South Carolina School of Law; B.S. & B.A., 2011, Coastal Carolina University. This article was written in conjunction with a discussion of this topic at the annual Law and Society Conference. This article could not have been possible without the helpful suggestions and edits provided by the following people: Professor F. Patrick Hubbard, Dr. Keith Guzik, James D. Perl, Stephen Sutherland, Alex Pate, James Johnson, Jarrod Kipp, and Darby Voisin. Most emphatically, I owe my greatest thanks to my fiancée Victoria Voisin and my brother Nicholas Gurney who were my springboards for many ideas and who edited the article countless times. And finally, I would like to thank the fine editors of the Albany Law Review whose work has greatly improved the quality of this article.

¹ ISAAC ASIMOV, *Runaround*, in I, ROBOT 30, 44–45 (1991) [hereinafter *Runaround*]. Later, Asimov created a fourth rule called the “Zeroth Law,” which states: “A robot may not injure humanity or, through inaction, allow humanity to come to harm.” F. Patrick Hubbard, “Do Androids Dream?: Personhood and Intelligent Artifacts, 83 TEMP. L. REV. 405, 463–64 (2011). Noah Goodall was the first to the author’s knowledge to replace “robot” with “autonomous vehicle” when discussing Asimov’s laws and autonomous vehicles. See Noah J. Goodall, *Ethical Decision Making During Automated Vehicle Crashes*, 2424 J. TRANSP. RES. BOARD 58, 61 (2014) [hereinafter Goodall, *Ethical Decision Making*].

² See *Runaround*, *supra* note 1, at 30, 44–45.

rules.³ These shortcomings arose due to the ambiguity of the rules.⁴

In addition to being impractical in Asimov's stories, these rules would be ineffective for programming autonomous vehicles.⁵ For example, how would an Asimovian autonomous vehicle address a situation where a truly unavoidable accident must occur and someone must be harmed? No matter how well autonomous vehicles are programmed, they will inevitably be involved in accidents.⁶ In

³ See Keith Abney, *Robotics, Ethical Theory, and Metaethics: A Guide for the Perplexed*, in ROBOT ETHICS: THE ETHICAL AND SOCIAL IMPLICATIONS OF ROBOTICS 35, 43 (Patrick Lin et al. eds., 2012). ("[I]n story after story, Asimov demonstrated that three simple, hierarchically arranged rules could lead to deadlocks when, for example, the robot received conflicting instructions from two people, or when protecting one person might cause harm to others."); Hubbard, *supra* note 1 at 465, 468 (discussing the shortcomings of Asimov's laws); Robin R. Murphy & David D. Woods, *Beyond Asimov: The Three Laws of Responsible Robots*, IEEE INTELLIGENT SYSTEMS, July–Aug. 2009, at 14, 14. ("Although the laws were simple and few, the stories attempted to demonstrate just how difficult they were to apply in various real-world situations. In most situations, although the robots usually behaved 'logically,' they often failed to do the 'right' thing."); Wendell Wallach, *The Challenge of Moral Machines*, PHIL. NOW, Mar.–Apr. 2009, at 6, 8 [hereinafter Wallach, *The Challenge of Moral Machines*] ("[I]n story after story Asimov demonstrated that even these three rather intuitive principles arranged hierarchically can lead to countless problems.").

⁴ Asimov himself stated in his introduction to Part II of *The Rest of Robots* that "[t]here was just enough ambiguity in the Three Laws to provide the conflicts and uncertainties required for new stories, and, to my great relief, it seemed always to be possible to think up a new angle out of the sixty-one words of the Three Laws." ISAAC ASIMOV, *THE REST OF THE ROBOTS* 43 (1964); see also Noah J. Goodall, *Machine Ethics and Automated Vehicles*, in ROAD VEHICLE AUTOMATION 93, 99 (Gereon Meyer & Sven Beiker eds., 2014) ("In Asimov's laws, an automated vehicle might avoid braking before a collision because this action would first give its occupants whiplash, thereby violating the first law prohibiting harm to humans.") [hereinafter Goodall, *Machine Ethics*]; Gabriel Hallevy, "I, Robot – I, Criminal"—When Science Fiction Becomes Reality: Legal Liability of AI Robots Committing Criminal Offenses, 22 SYRACUSE SCI. & TECH. L. REP. 1, 1–2 (2010) ("These three fundamental laws are obviously contradictory. What if a man orders a robot to hurt another person for the own good of the other person? What if the robot is in police service and the commander of the mission orders it to arrest a suspect and the suspect resists arrest? Or what if the robot is in medical service and is ordered to perform a surgical procedure on a patient, the patient objects, but the medical doctor insists that the procedure is for the patient's own good, and repeats the order to the robot?").

⁵ See Murphy & Woods, *supra* note 3, at 15. (discussing the shortcomings of applying Asimov's laws to modern robots).

⁶ See, e.g., Goodall, *Ethical Decision Making*, *supra* note 1, at 59 ("It is not difficult to imagine a scenario in which a crash is unavoidable, even for an automated vehicle with complete knowledge of its world and negligible reaction time."); David C. Vladeck, *Machines Without Principals: Liability Rules and Artificial Intelligence*, 89 WASH. L. REV. 117, 126–27 (2014) ("Like human drivers, the machines that drive Google cars will on occasion encounter unexpected events that call for snap judgments: a child darting in front of a car; a tree limb crashing down just a few yards ahead; a car running a red light; or a patch of black ice that is undetectable on a moonless night. There are liability rules that come into play when, as a consequence of any of these unexpected events, there is injury to humans or damage to property. No matter how well-designed and programmed self-driving cars are, factors beyond the machine's control virtually guarantee that at some point the car will have an accident that will cause injury of some kind, and will act in ways that are not necessarily ordained by their programming."); Patrick Lin, *The Robot Car of Tomorrow May Just Be Programmed to Hit You*, WIRED (May 6, 2014) [hereinafter *The Robot Car of Tomorrow*]

Asimov's short story, titled "Liar," a robot named Herbie had the ability to read people's minds.⁷ Herbie read two characters' minds and told each what they wanted to hear—lies—so that he would not cause "emotional harm," which Herbie considered to be in violation of the first rule.⁸ Based on what Herbie told them, the characters changed their behavior; eventually, a disagreement ensued.⁹ The characters confronted Herbie to sort out their mess.¹⁰ Herbie, realizing that he could not answer the question without causing harm to someone, broke down and stopped working.¹¹

An autonomous vehicle that breaks down when it encounters a truly unavoidable accident would be impractical and dangerous.¹² Shutting down would not prevent the accident, and would most likely aggrandize it, and thus, would still violate the first rule. For that reason, and many others, Asimov's laws are impractical for addressing the ethical issues created by autonomous vehicles.¹³

Instead of the autonomous technology shutting down when faced with an unavoidable accident, society will want the autonomous vehicle to minimize the amount of harm that results from such an accident, regardless of who is at fault.¹⁴ An algorithm writer can minimize harm that results from an accident through use of a "crash-optimization algorithm," which is the method by which an

<http://www.wired.com/2014/05/the-robot-car-of-tomorrow-might-just-be-programmed-to-hit-you/> ("Some road accidents are unavoidable, and even autonomous cars can't escape that fate.").

⁷ See ISAAC ASIMOV, *Liar*, in I, ROBOT 111, 127 (1991).

⁸ See *id.* at 129–31.

⁹ See *id.* at 121–35.

¹⁰ See *id.* at 129–30.

¹¹ See *id.* at 134.

¹² See WENDELL WALLACH & COLIN ALLEN, MORAL MACHINES: TEACHING ROBOTS RIGHT FROM WRONG 92, 93 (2009) (suggesting that a robot that breaks down to prevent harm to humans could, conversely, cause even more harm).

¹³ But see Raul Rojas, *I, Car: The Four Laws of Robotic Cars*, FREIE UNIVERSITÄT BERLIN, http://www.inf.fu-berlin.de/inst/ag-ki/rojas_home/documents/tutorials/I-Car-Laws.pdf (last visited Jan. 09, 2016) (noting how there only needs to be the addition of a new rule on top of Asimov's original three). Rojas suggests four laws of robotic cars:

1. A car may not injure a human being or, through inaction, allow a human being to come to harm.
2. A car must obey the traffic rules, except when they would conflict with the First Law.
3. A car must obey the orders given to it by human beings, except where such orders would conflict with the First or Second Laws.
4. A car must protect its own existence as long as such protection does not conflict with the First, Second or Third Laws.

Id. As indicated, situations will occur where injury to a human being must result, and thus, these rules do not resolve that problem. Indeed, Rojas himself acknowledged that these four rules have flaws. *Id.*

¹⁴ See Patrick Lin, *Why Ethics Matters for Autonomous Cars*, in AUTONOMES FAHREN 69, 72 (2015) [hereinafter *Why Ethics Matters*] (stating that society will want autonomous vehicles to minimize harm when faced with an unavoidable accident).

autonomous vehicle determines who or what to hit.¹⁵ The nature of the decisions made by the algorithm writer in creating the crash-optimization algorithm implicates serious ethical and legal questions.

This article examines those ethical and legal questions. Part II begins by providing background information on autonomous vehicles, and the benefits that those vehicles are projected to bring to society. Part III introduces six moral dilemmas as a lens through which to examine the ethical and legal questions arising out of a crash-optimization algorithm. In Part IV, the article provides an overview of Utilitarian and Kantian ethics, as well as the application of both ethical theories to autonomous vehicles. Part V examines tort and criminal law issues relating to the crash-optimization algorithms. Part VI initially addresses whether these decisions should even be made by robotic cars. After concluding that the decisions should be made by the autonomous vehicles, Part VI examines who—the car owner, the car manufacturer, or the government—should make that decision. Finally, the article provides a legal framework for the application of criminal and tort law to accidents involving the use of crash-optimization algorithms.

II. OVERVIEW OF AUTONOMOUS VEHICLES

A. Background

Since at least 1939, when General Motors introduced the Futurama exhibit at the World's Fair in New York, generation after generation has awaited the arrival of autonomous vehicles.¹⁶ However, it was not until recent years that automated technology has become more advanced.¹⁷ Today's vehicles are equipped with such

¹⁵ *Id.* ("Optimizing crashes means to choose the course of action that will likely lead to the least amount of harm."); see also *Robot Ethics: Morals and the Machine*, ECONOMIST (June 2, 2012), <http://www.economist.com/node/21556234> ("[A]utonomous machines are bound to end up making life-or-death decisions in unpredictable situations."); *The Robot Car of Tomorrow*, *supra* note 6 ("Some road accidents are unavoidable, and even autonomous cars can't escape that fate."); 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/> ("We are moving closer to having driverless cars on roads everywhere, and naturally, people are starting to wonder what kinds of ethical challenges driverless cars will pose. One of those challenges is choosing *how* a driverless car should react when faced with an unavoidable crash scenario.").

¹⁶ See Burkhard Bilger, *Auto Correct: Has the Self-Driving Car at Last Arrived?*, NEW YORKER, Nov. 25, 2013, at 96, 96, 98.

¹⁷ Alex Davies, *The Sneaky Way Automakers Are Getting Us to Accept Self-Driving Cars*, WIRED (May 30, 2014), <http://www.wired.com/2014/05/automakers-self-driving-cars/>

autonomous technology as automatic lane-keeping, automatic braking, adaptive cruise control, traffic jam assist, and automated parallel-parking.¹⁸ Mercedes-Benz's E- and S-Class models utilize "Stop&Go Pilot," which navigates the vehicle in traffic jams.¹⁹ The 2016 Cadillacs and Tesla's new Model S will be capable of driving themselves on highways.²⁰

The race to create automated vehicles accelerated after the Defense Advanced Research Projects Agency ("DARPA") held its first "Grand Challenge" in 2004.²¹ The first Grand Challenge was a 142-mile autonomous vehicle race in the Mojave Desert.²² Unfortunately, all of the cars failed shortly after the race started.²³ The following year, four of the twenty-three cars completed a 132-mile course in the second Grand Challenge.²⁴ Since then, all major car companies, (and Google) have been in the process of developing technology to allow for self-driving vehicles.²⁵ Although no autonomous vehicle is on the

[hereinafter *The Sneaky Way*].

¹⁸ Julie Goodrich, Comment, *Driving Miss Daisy: An Autonomous Chauffeur System*, 51 HOUS. L. REV. 265, 267–68 (2013) (describing various automated technologies utilized in vehicles); accord Andrew P. Garza, Note, "Look Ma, No Hands!": Wrinkles and Wrecks in the Age of Autonomous Vehicles, 46 NEW ENG. L. REV. 581, 584–85, 586 (2012); see also *The Sneaky Way*, *supra* note 17 ("Automatic braking is just one of many autonomous technologies the big car companies have introduced in their vehicles in recent years. Features like adaptive cruise control, lane keeping assist, pedestrian recognition, and parking assist are increasingly common."); Aaron M. Kessler, *Technology Takes the Wheel*, N.Y. TIMES (Oct. 5, 2014), <http://www.nytimes.com/2014/10/06/business/technology-takes-the-wheel.html>; Joseph B. White, *How Do You Insure a Driverless Car?*, WALL ST. J. (Aug. 7, 2014), <http://www.wsj.com/articles/driverless-cars-edge-onto-roadways-1407432414> ("[A]n increasing number of car makers will offer 'traffic jam assist' systems that take over braking, steering and acceleration for vehicles inching along in low-speed traffic.").

¹⁹ Alex Davies, *The Mercedes Robo-Car That Made Me Want to Stop Driving*, WIRED (Mar. 23, 2015), <http://www.wired.com/2015/03/mercedes-benz-f-015-autonomous-car/> [hereinafter *Mercedes Robo-Car*].

²⁰ Jeff Bennett & Joseph B. White, *GM Expects to Offer Hands-Free Driving by 2016*, WALL ST. J. (Sept. 7, 2014), <http://www.wsj.com/articles/gm-expects-to-offer-hands-free-driving-by-2016-1410124540> (discussing Cadillac); Alex Davies, *Elon Wants to Make Your Tesla Drive Itself. Is That Legal?*, WIRED (Mar. 19, 2015), <http://www.wired.com/2015/03/elon-wants-make-tesla-drive-legal/> [hereinafter *Elon*] (discussing Tesla).

²¹ Matthew Michaels Moore & Beverly Lu, *Autonomous Vehicles for Personal Transport: A Technology Assessment 1* (June 2, 2011) (unpublished article) (on file with the California Institute of Technology) ("A large portion of the recent progress in autonomous vehicle technology can be directly attributed to two competitions staged by the Defense Advanced Research Projects Agency (DARPA).").

²² Sebastian Thrun, *Toward Robotic Cars*, COMMS. OF THE ACM, Apr. 2010, at 99, 99.

²³ *Id.*

²⁴ *Id.*

²⁵ ERNST & YOUNG, *DEPLOYING AUTONOMOUS VEHICLES: COMMERCIAL CONSIDERATIONS AND URBAN MOBILITY SCENARIOS 2* (2014), [http://www.ey.com/Publication/vwLUAssets/EY-Deploying-autonomous-vehicles-30May14/\\$FILE/EY-Deploying-autonomous-vehicles-30May14.pdf](http://www.ey.com/Publication/vwLUAssets/EY-Deploying-autonomous-vehicles-30May14/$FILE/EY-Deploying-autonomous-vehicles-30May14.pdf) ("[A]lmost every major vehicle manufacturer (VM), supplier and technology company has announced projects or collaborations around the autonomous vehicles (AVs)

road yet, it is projected that these vehicles will be operated by a complex computer system, with the use of radar, laser, lidar, ultrasonic sensors, video cameras, global positioning systems, and maps.²⁶ This technology allows the vehicle to safely operate by constantly watching the road and obstacles in its view.²⁷ Most automakers intend to keep the human operator in the loop,²⁸ at least at first.²⁹

Autonomous vehicles have demonstrated both safety and efficiency on the roadway. Google's autonomous vehicles have travelled over two million miles without causing an accident;³⁰ an autonomous vehicle created by the automotive supplier Delphi drove from San

theme."); DANIEL J. FAGNANT & KARA KOCKELMAN, ENO CTR. FOR TRANSP., PREPARING A NATION FOR AUTONOMOUS VEHICLES: OPPORTUNITIES, BARRIERS AND POLICY RECOMMENDATIONS 2 (2013), <https://www.enotrans.org/wp-content/uploads/wpsc/downloadables/AV-paper.pdf> ("[Google] and numerous manufacturers—including Audi, BMW, Cadillac, Ford, GM, Mercedes-Benz, Nissan, Toyota, Volkswagen, and Volvo—have begun testing driverless systems."); Alex Davies, *This is Big: A Robo-Car Just Drove Across the Country*, WIRED (Apr. 3, 2015), <http://www.wired.com/2015/04/delphi-autonomous-car-cross-country/> [hereinafter *This is Big*] ("[M]ost of the world's major automakers are working on autonomous technology, [even] with Audi, Mercedes-Benz, Nissan, and Volvo leading the pack[,] Google may be more advanced than anyone.").

²⁶ Kyle Colonna, *Autonomous Cars and Tort Liability*, 4 CASE W. RES. J.L. TECH. & INTERNET 81, 86–87 (2012); Sophia H. Duffy & Jamie Patrick Hopkins, *Sit, Stay, Drive: The Future of Autonomous Car Liability*, 16 SMU SCI. & TECH. L. REV. 453, 455 (2013); see also Paul Stenquist, *Nissan Announces Plans to Release Driverless Cars by 2020*, N.Y. TIMES: WHEELS (Aug. 29, 2013, 6:00 AM), <http://wheels.blogs.nytimes.com/2013/08/29/nissan-announces-plans-to-release-driverless-cars-by-2020/> ("A host of advanced equipment is needed for autonomous operation, including cameras that can see the area surrounding the vehicle; radar sensors that measure distance; laser scanners that detect the shape of objects; a global-positioning sensor that locates the vehicle; advanced computer systems that apply artificial intelligence to that data and make driving decisions; and a variety of actuators that can execute driving maneuvers while compensating for less than ideal conditions.").

²⁷ Vladeck, *supra* note 6, at 126.

²⁸ See Stenquist, *supra* note 26. State laws specifically addressing autonomous vehicles also require the autonomous vehicle to have an easy method for the operator to retake control of the vehicle. See, e.g., CAL. VEH. CODE § 38750(c)(1)(A), (D) (West 2015); D.C. CODE § 50-2352(1) (2015); FLA. STAT. ANN. § 319.145(1)(b) (West 2015); NEV. ADMIN. CODE § 482A.190(2)(b) (2015).

²⁹ Although Google originally intended to keep humans in the loop, it has now decided to remove humans from the loop because it realized that they do not make for "trustworthy" co-pilots. Douglas MacMillan & Rolfe Winkler, *Google's Prototype for Autonomous Driving Has No Steering Wheel*, WALL ST. J.: DIGITS (May 27, 2014, 11:55 PM), http://blogs.wsj.com/digits/2014/05/27/googles-prototype-for-autonomous-driving-has-no-steering-wheel; Tom Simonite, *Lazy Humans Shaped Google's New Autonomous Car*, MIT TECH. REV. (May 30, 2014), <http://www.technologyreview.com/news/527756/lazy-humans-shaped-googles-new-autonomous-car/>. Thus, Google's autonomous vehicles will not have a steering wheel, accelerator, or a brake pedal. See Lee Gomes, *Hidden Obstacles for Google's Self-Driving Cars*, MIT TECH. REV. (Aug. 28, 2014), <http://www.technologyreview.com/news/530276/hidden-obstacles-for-googles-self-driving-cars/>.

³⁰ See Keith Naughton, *Humans Are Slamming Into Driverless Cars and Exposing a Key Flaw*, BLOOMBERG (Dec. 17, 2015), <http://www.bloomberg.com/news/articles/2015-12-18/humans-are-slamming-into-driverless-cars-and-exposing-a-key-flaw>. The cars have been involved in seventeen accidents, all of which have been caused by other drivers. *Id.*

Francisco to New York City within nine days.³¹ These trends have led many experts to predict that fully autonomous vehicles will be available within decades,³² and that they will be commonplace by 2040.³³

The federal government has also begun researching autonomous vehicles. The National Highway Traffic Safety Administration (“NHTSA”) released a preliminary statement of policy regarding these vehicles.³⁴ In that statement, the NHTSA outlined its five levels of automation:

Level 0 – No-Automation: The driver is in complete and sole control of the primary vehicle controls (brake, steering, throttle, and motive power) at all times, and is solely responsible for monitoring the roadway and for safe operation of all vehicle controls. . . .

Level 1 – Function-specific Automation: Automation at this level involves one or more specific control functions; if multiple functions are automated, they operate independently from each other. The driver has overall control, and is solely responsible for safe operation

Level 2 – Combined Function Automation: This level involves automation of at least two primary control functions designed to work in unison to relieve the driver of control of those functions. . . .

Level 3 – Limited Self-Driving Automation: Vehicles at this level of automation enable the driver to cede full control of all safety-critical functions under certain traffic or environmental conditions and in those conditions to rely heavily on the vehicle to monitor for changes in those conditions requiring transition back to driver control. . . .

Level 4 – Full Self-Driving Automation (Level 4): The vehicle is designed to perform all safety-critical driving functions and monitor roadway conditions for an entire trip.³⁵

Current automated vehicle technology is between Levels 2 and 3.³⁶

³¹ *This is Big*, *supra* note 25. Although the car did 99% of the driving, it let its operator drive on city streets. *Id.*

³² *See, e.g., Mercedes Robo-Car*, *supra* note 19 (noting that Mercedes predicts autonomous vehicles will be available in fifteen years); Stenquist, *supra* note 26 (projecting automated vehicle production by 2020).

³³ *This is Big*, *supra* note 25.

³⁴ NAT’L HIGHWAY TRAFFIC SAFETY ADMIN., U.S. DEP’T OF TRANSP., PRELIMINARY STATEMENT OF POLICY CONCERNING AUTOMATED VEHICLES (2013) [hereinafter NHTSA].

³⁵ *Id.* at 4–5.

³⁶ Irving Wladawsky-Berger, *When Can We Expect Truly Autonomous Vehicles?*, WALL ST.

Because this technology is not at Level 3 or Level 4, the NHTSA believes it is premature to issue regulations.³⁷ The NHTSA signaled its encouragement for the innovation of autonomous vehicles, but it cautioned that they should not be used on public roads except for testing purposes.³⁸

Some states and the District of Columbia are foreshadowing the arrival of Level 3 and Level 4 autonomous vehicles onto their roadways and have already enacted autonomous vehicle laws. Thus far, California, Florida, Michigan, Nevada, and the District of Columbia have enacted autonomous vehicle laws that explicitly allow for testing of autonomous vehicles on their roads.³⁹ None of these laws, however, address crash-optimization algorithms or, more generally, tort liability for accidents caused by autonomous vehicles.⁴⁰

In addition, foreign countries are enacting autonomous vehicle laws and fostering the growth of such technology.⁴¹ For example, the United Kingdom is on the cutting edge of the autonomous vehicle movement.⁴² Starting this year, the United Kingdom is funding autonomous shuttles in the cities of Greenwich, Bristol, Milton Keynes, and Coventry.⁴³ The shuttle resembles an elongated golf cart.⁴⁴ A two-seater vehicle called a “pod” will be operated on the

J.: CIO REP. (Oct. 31, 2014, 10:39 AM), <http://blogs.wsj.com/cio/2014/10/31/when-can-we-expect-truly-autonomous-vehicles/>.

³⁷ NHTSA, *supra* note 34, at 10.

³⁸ *Id.*

³⁹ See CAL. VEH. CODE § 38750(b) (West 2015); D.C. CODE § 50-2352 (2015); FLA. STAT. ANN. § 316.86(1) (West 2015); MICH. COMP. LAWS ANN. §§ 257.663, 257.665(1) (West 2014); NEV. REV. STAT. ANN. § 482A.080(2) (LexisNexis 2014).

⁴⁰ The one exception is a Michigan statute that limits liability for the manufacturer of a vehicle when another person adds autonomous technology to that manufacturer's vehicle. See MICH. COMP. LAWS ANN. § 600.2949b(1) (West 2015).

⁴¹ See, e.g., Kate Connolly, *Germany Creates Laws for Driverless Cars*, GUARDIAN (Feb. 1, 2015), <http://www.theguardian.com/technology/2015/feb/01/germany-laws-driverless-cars-auto-bahns-google-industry> (discussing legal guidelines for driverless cars in Germany); Peter Murray, *China Gains on Google in Driverless Car Race with 177 Mile Road Trip*, SINGULARITYHUB (Aug. 7, 2011), <http://singularityhub.com/2011/08/07/china-gains-on-google-in-driverless-car-race-with-177-mile-road-trip-video/> (discussing a driverless vehicle in China, and its 177 mile journey on real roads); *Japanese Companies, Ministries Come Together to Hasten Development*, NIKKEI ASIAN REV. (Feb. 26, 2015), <http://asia.nikkei.com/Business/Trends/Japanese-companies-ministries-come-together-to-hasten-development> (discussing measures being taken in Japan to support self-driving cars); Gwyn Topham, *Driverless Cars Set to Roll Out for Trials on UK Roads*, GUARDIAN (Feb. 10, 2015), <http://www.theguardian.com/technology/2015/feb/11/driverless-cars-roll-out-trials-uk-roads> (explaining driverless cars in the United Kingdom).

⁴² Topham, *supra* note 41.

⁴³ *Id.*

⁴⁴ *Id.*

streets of Milton Keynes.⁴⁵

B. Benefits of Autonomous Vehicles

A major reason why there is so much interest in autonomous vehicle development by car manufacturers—and so much support for this development from governments worldwide—is the societal benefits that autonomous vehicles are projected to have. Approximately 1.24 million people die annually worldwide due to accidents on roadways.⁴⁶ In 2013, 32,719 Americans died and 2,313,000 American were injured in car crashes; a total of 5,657,000 accidents occurred in the United States that year.⁴⁷ To put the amount of Americans who die on the roadways into perspective, more Americans died from motor vehicle accidents during the United States' involvement in World Wars I and II, the Korean War, and the Vietnam War than Americans who died defending the country in those same years.⁴⁸ This has led one commenter to suggest that self-driving cars will save more lives than world peace.⁴⁹

Most experts believe that autonomous vehicles will prevent accidents because ninety percent of accidents are caused by driver error.⁵⁰ Unlike the human driver, the autonomous vehicle “sees” everything in the vicinity; reacts at speeds humans cannot match; and constantly checks the performance of every component in the vehicle to ensure that it is functioning properly.”⁵¹ The autonomous

⁴⁵ *See id.*

⁴⁶ WORLD HEALTH ORG., GLOBAL STATUS REPORT ON ROAD SAFETY 2013 1 (2013).

⁴⁷ NAT'L HIGHWAY TRAFFIC SAFETY ADMIN., U.S. DEP'T OF TRANSP., *2013 Crash Overview* (Jan. 15, 2015), <http://www-nrd.nhtsa.dot.gov/Pubs/13WPPP.pdf> [hereinafter *2013 Crash Overview*].

⁴⁸ Matt McFarland, *How Self-Driving Cars Would Benefit Americans More Than World Peace*, WASH. POST (Feb. 10, 2015), <http://www.washingtonpost.com/blogs/innovations/wp/2015/02/10/how-self-driving-cars-would-benefit-americans-more-than-world-peace/>. In this article, McFarland noted that the war deaths from World War I (116,516) and World War II (405,399) were more than the automotive deaths (20,020 and 137,826, respectively) from those eras, but that the deaths due to car crashes during the Korean War and Vietnam War (140,773 and 558,506, respectively) substantially outnumbered the number of Americans who died bravely defending our country in those wars (36,516 and 58,220, respectively). *Id.* McFarland summarized that during those time periods a total of 857,125 Americans died from car accidents, while 616,651 Americans died defending our country. *Id.*

⁴⁹ *See id.*

⁵⁰ Sven A. Beiker, *Legal Aspects of Autonomous Driving: The Need for a Legal Infrastructure That Permits Autonomous Driving in Public to Maximize Safety and Consumer Benefit*, 52 SANTA CLARA L. REV. 1145, 1149 (2012); Noah J. Goodall, Presentation at the 21st World Congress on Intelligent Transport Systems: Vehicle Automation and the Duty to Act 1 (Sept. 2014) (stating that about ninety-three percent of accidents are caused by human error) [hereinafter Goodall, Presentation].

⁵¹ Vladeck, *supra* note 6, at 126.

vehicle also does not drink alcohol; it does not get drowsy; and it does not make phone calls, text, eat, or engage in any other activity that distracts the human driver.⁵² Therefore, it is likely that autonomous vehicles will greatly reduce the number of accidents. One study has shown that when autonomous vehicles reach ten percent penetration in the marketplace, 1,100 fewer people will die and 211,000 fewer accidents will occur on American roadways.⁵³ At ninety percent penetration, 21,700 fewer people will die and 4,220,000 fewer crashes will occur annually.⁵⁴ The prevention of car accidents will also reduce other societal costs—such as hospital stays, days of work missed, and the emotional toll accidents have on families.⁵⁵

Drivers in the United States spend an average of seventy-five billion hours per year commuting,⁵⁶ and the average driver spends fifty-one minutes commuting to work daily.⁵⁷ Over 3.5 million people commute at least ninety minutes just to get to work.⁵⁸ Study after study has shown that daily commutes make people unhappy and have other negative effects on their lives.⁵⁹ Autonomous vehicles should be able to alleviate much unhappiness that results from a

⁵² See, John Markoff, *Google Cars Drive Themselves, in Traffic*, N.Y. TIMES (Oct. 9, 2010), <http://www.nytimes.com/2010/10/10/science/10google.html> (“Robot drivers react faster than humans, have 360-degree perception and do not get distracted, sleepy or intoxicated.”); Claire Cain Miller & Matthew L. Wald, *Self-Driving Cars for Testing are Supported by U.S.*, N.Y. TIMES (May 30, 2013), <http://www.nytimes.com/2013/05/31/technology/self-driving-cars-for-testing-are-supported-by-us.html?pagewanted=all> (“Autonomous cars could increase safety because they are not subject to human error like disobeying traffic laws and falling asleep at the wheel.”).

⁵³ FAGNANT & KOCKELMAN, *supra* note 25, at 8 tbl.2.

⁵⁴ *Id.*

⁵⁵ NHTSA, *supra* note 34, at 1.

⁵⁶ MORGAN STANLEY, AUTONOMOUS CARS: SELF-DRIVING THE NEW AUTO INDUSTRY PARADIGM 16 (2013), <http://orfe.princeton.edu/~alaink/SmartDrivingCars/PDFs/Nov2013MORGAN-STANLEY-BLUE-PAPER-AUTONOMOUS-CARS%EF%BC%9A-SELF-DRIVING-THE-NEW-AUTO-INDUSTRY-PARADIGM.pdf>.

⁵⁷ See Larry Copeland, *Americans' Commutes Aren't Getting Longer*, USA TODAY (Mar. 5, 2013), <http://www.usatoday.com/story/news/nation/2013/03/05/americans-commutes-not-getting-longer/1963409>.

⁵⁸ Nick Paumgarten, *There and Back Again: The Soul of the Commuter*, NEW YORKER (Apr. 16, 2007), <http://www.newyorker.com/magazine/2007/04/16/there-and-back-again>.

⁵⁹ See Daniel Kahneman & Alan B. Krueger, *Developments in the Measurement of Subjective Well-Being*, 20 J. ECON. PERSP. 3, 13 (2006) (finding that the morning commute makes people the least happy); Erika Sandow, *Til Work Do Us Part: The Social Fallacy of Long-Distance Commuting*, 51 URB. STUD. 526, 529 (2014) (noting a finding that people who commute at least an hour one way to work would need a pay raise of forty percent to have the same happiness as a non-commuter); Alois Stutzer & Bruno S. Frey, *Stress that Doesn't Pay: The Commuting Paradox*, 110 SCANDINAVIAN J. ECON. 339, 363 (2008) (“[There is] a large negative effect of commuting time on people’s satisfaction with life.”); see also Annie Lowrey, *Your Commute Is Killing You*, SLATE (May 26, 2011), http://www.slate.com/articles/business/moneybox/2011/05/your_commute_is_killing_you.single.html (summarizing studies on the effects of commuting).

daily commute because people will no longer need to pay attention to the road for their vehicles to safely operate, and they will be able to engage in activities during the commute, such as reading, sleeping, or working.⁶⁰ Indeed, a main reason that people will purchase autonomous vehicles is so they can engage in other activities.

Additionally, people who cannot safely operate an automobile due to age or disability will have an incentive to purchase automated vehicles. Currently, people who are under a certain age, or physically disabled are prohibited from driving by law.⁶¹ Level 4 autonomous vehicles can drive safely without human intervention;⁶² therefore, these vehicles have the potential to provide people who cannot currently drive with an opportunity to increase their mobility and independence, which should lead to happier lives.⁶³

Autonomous vehicles will also reduce congestion and fuel consumption.⁶⁴ Currently, “[t]raffic jams account for 3.7 billion wasted hours of human time and 2.3 billion wasted gallons of fuel.”⁶⁵ Autonomous vehicles will reduce traffic congestion for a variety of

⁶⁰ See JAMES M. ANDERSON ET AL., RAND CORP., AUTONOMOUS VEHICLE TECHNOLOGY: A GUIDE FOR POLICYMAKERS, 18 (2014), http://www.rand.org/content/dam/rand/pubs/research_reports/RR400/RR443-1/RAND_RR443-1.pdf (“[Autonomous vehicles will] free drivers to engage in other productive or enjoyable activities—working, reading, watching a movie, or even sleeping—during a trip, thus reducing the opportunity cost of time spent in the car.”); MORGAN STANLEY, *supra* note 56, at 16 (“US drivers spend an average of 75 billion hours each year on the road, which can now be put to good use. Whether people choose to spend this time eating, sleeping, watching TV, reading the newspaper, working, or simply conversing, it should result in significantly de-stressing the average commute and life in general.”).

⁶¹ See, e.g., N.Y. VEH. & TRAF. LAW § 502(2) (McKinney 2015) (stating the age requirements for the different classes of licenses in New York); N.Y. VEH. & TRAF. LAW § 509(9) (McKinney 2015) (stating that whenever a person becomes disabled they cannot operate a motor vehicle until they notify the commissioner of the DMV); N.Y. COMP. CODES R. & REGS. tit. 15, § 6.10(b) (2015) (describing physical disqualifications for legally operating a bus); Beiker, *supra* note 50, at 1151–52 (describing how autonomous driving assistance technology would help those that otherwise would have trouble driving, such as young people and people with disabilities).

⁶² For a discussion on the levels of autonomous technology, see *supra* note 35 and accompanying text.

⁶³ See, e.g., ANDERSON ET AL., *supra* note 60, at 16–17 (“Level 4 vehicles could substantially increase access and mobility across a range of populations currently unable or not permitted to use conventional automobiles. These include the disabled, older citizens, and children under the age of 16.”); Beiker, *supra* note 50, at 1151–52 (indicating that autonomous vehicles could provide freedom to adolescents, elderly, and persons suffering from disabilities); Dana M. Mele, *The Quasi-Autonomous Car as an Assistive Device for Blind Drivers: Overcoming Liability and Regulatory Barriers*, 28 SYRACUSE J. SCI. & TECH. L. REP. 26, 28 (2013) (“[Autonomous vehicles] will contribute to goals of independence and autonomy for individuals with disabilities.”); Bryant Walker Smith, *Managing Autonomous Transportation Demand*, 52 SANTA CLARA L. REV. 1401, 1409 (2012) (“Self-driving cars that do not need human drivers or monitors may substantially increase mobility for those who cannot (legally) drive themselves because of youth, age, disability, or incapacitation.”).

⁶⁴ FAGNANT & KOCKELMAN, *supra* note 25, at 4.

⁶⁵ Thrun, *supra* note 22, at 99.

reasons. Approximately twenty-five percent of traffic congestion is due to traffic accidents.⁶⁶ As indicated above, autonomous vehicles should greatly reduce the number of accidents on the road.⁶⁷ Autonomous vehicles will also be able to coordinate and anticipate traffic more precisely than human drivers, which will harmonize traffic flow.⁶⁸ In addition, the ability to smoothly accelerate and brake will decrease traffic congestion and enable vehicle platooning.⁶⁹

Although vehicles are one of the costliest assets that people purchase,⁷⁰ car owners only utilize their vehicles for about four percent of their lifetime.⁷¹ Car-sharing services, such as Uber and Lyft, are becoming commonplace, and it is likely that autonomous vehicles will greatly increase the abilities of car sharing services in the United States.⁷² Autonomous vehicles can provide an “unprecedented level of convenience” for car sharing because they can drive passengers and return to their charge stations.⁷³ If an autonomous vehicle is shared such that it replaces four vehicles in the United States, such vehicle-sharing may lead to about \$1.8 trillion in savings annually.⁷⁴

Even if such car-sharing does not materialize, the benefits of autonomous vehicles will result in enormous economic savings for society. Morgan Stanley has predicted that, upon full penetration of autonomous vehicles into the market, such vehicles “can contribute \$1.3 trillion in annual savings to the US economy alone, with global

⁶⁶ FAGNANT & KOCKELMAN, *supra* note 25, at 5.

⁶⁷ *See id.* at 4.

⁶⁸ Beiker, *supra* note 50, at 1150; *see also* Farhad Manjoo, *The Future Could Work, if We Let It*, N.Y. TIMES (Aug. 27, 2014), http://www.nytimes.com/2014/08/28/technology/personaltech/technology-and-the-human-factor-the-future-could-work-if-we-let-it.html?_r=0 (“Autonomous vehicles . . . will be able to pack roads more efficiently. We could get eight times as many cars on a freeway without slowing down, letting us get around faster and, in time, build and maintain fewer roads.”).

⁶⁹ FAGNANT & KOCKELMAN, *supra* note 25, at 5; Robert B. Kelly & Mark D. Johnson, *Defining a Stable, Protected and Secure Spectrum Environment for Autonomous Vehicles*, 52 SANTA CLARA L. REV. 1271, 1277–78 (2012) (discussing that a “vehicle platoon” occurs when the autonomous vehicles “speak” wirelessly between each other to form a train, which will increase roadway efficiency); Smith, *supra* note 63, at 1412.

⁷⁰ Thrun, *supra* note 22, at 99.

⁷¹ *Id.* at 105.

⁷² *See, e.g.*, KEVIN SPIESER ET AL., *Toward a Systematic Approach to the Design and Evaluation of Automated Mobility-on-Demand Systems: A Case Study in Singapore*, in ROAD VEHICLE AUTOMATION 229, 243 (Gereon Meyer & Sven Beiker eds., 2014); Thrun, *supra* note 22, at 105; Emilio Frazzoli, *Can We Put a Price on Autonomous Driving?*, MIT TECH. REV. (Mar. 18, 2014), <http://www.technologyreview.com/view/525591/can-we-put-a-price-on-autonomous-driving/> (discussing the benefits of car sharing).

⁷³ Frazzoli, *supra* note 72.

⁷⁴ *Id.*

savings estimated at over \$5.6 trillion.”⁷⁵ The \$1.3 trillion represents \$507 billion in productivity gains, \$488 billion from accident avoidance, \$158 billion in fuel savings, \$138 billion in productivity gains from congestion avoidance, and \$11 billion from fuel savings due to congestion avoidance.⁷⁶ Therefore, the proliferation of autonomous vehicles will greatly benefit society, and their arrival into society should be welcomed.

III. MORAL DILEMMAS

This Part introduces six moral dilemmas as a method of analyzing crash-optimization algorithms. Each moral dilemma includes discussions on the ethical and legal issues that relate to the use of a crash-optimization algorithm.

A. *The Shopping Cart Problem*

After pulling into the parking lot of a grocery store, an autonomous vehicle’s brakes stop working. Directly in front of the autonomous vehicle is a mother pushing a baby in a baby carriage. To its left is an overloaded shopping cart, and to its right is the grocery store. Assuming that any of the choices would be capable of stopping the autonomous vehicle, what should the autonomous vehicle do?⁷⁷

This problem provides an example as to which choice not to make. No one in this situation would hit the baby carriage, unless the person knew that the baby carriage was empty. And we would expect, if not require, the autonomous vehicle to make that same choice. The decision of what the autonomous vehicle hits depends on the vehicle’s crash-optimization algorithm. One potentially popular choice of programming a crash-optimization algorithm would be to have the vehicle protect itself and its occupants. Thus, the rule for the algorithm—hereinafter, the “radical self-preservation” algorithm—would be that the vehicle should mitigate accidents by focusing solely on the damage to the vehicle and the well-being of the vehicle’s occupants.

⁷⁵ MORGAN STANLEY, *supra* note 56, at 1.

⁷⁶ *Id.* at 9.

⁷⁷ The idea of the autonomous vehicle striking either the shopping cart or the baby carriage has been raised by others. See, e.g., Miller & Wald, *supra* note 52 (“The first time that a driverless vehicle swerves to avoid a shopping cart and hits a stroller, someone’s going to write, ‘robot car kills baby to save groceries,’” [Ryan Calo] said. ‘It’s those kinds of reasons you want to make sure this stuff is fully tested.’”); Vivek Wadhwa, *Move Over, Humans, the Robocars Are Coming*, WASH. POST. (Oct. 14, 2014), <http://www.washingtonpost.com/blogs/innovations/wp/2014/10/14/move-over-humans-the-robocars-are-coming/>.

Under the radical self-preservation algorithm, the autonomous vehicle would not turn right into the grocery store. Hitting the grocery store would probably destroy the vehicle and seriously harm the occupants. Therefore, for the radical self-preservation autonomous vehicle, the decision is between turning left and hitting the full shopping cart or staying the course and hitting the baby carriage. Turning left to hit the shopping cart would result in damage to the autonomous vehicle. Most shopping carts are metal, and because this cart is full of groceries, it will be less forgiving than an empty shopping cart. Staying the course to hit the baby carriage is also likely to damage the autonomous vehicle, but it is clear that this would result in less damage to the car than hitting the shopping cart.⁷⁸ A baby carriage is mostly plastic and metal with a canvas covering. Therefore, the radical self-preservation autonomous vehicle would likely strike the baby carriage and kill the child, which is a “morally unacceptable result.”⁷⁹

The impact of an autonomous vehicle choosing to kill a baby instead of destroying groceries cannot be understated. The child’s family would be emotionally devastated, and would likely never recover from the trauma. The owner and occupants of the autonomous vehicle would also be seriously emotionally impacted. In addition, such a result could lead to the demise of autonomous vehicles.⁸⁰ Newspapers and 24/7 cable news would run the headline: “Robot Car Kills Baby to Avoid Groceries.”⁸¹ Public opinion of autonomous vehicles would turn negative, and movements to ban autonomous vehicles would form.⁸² Thus, the autonomous vehicle’s

⁷⁸ See *Why Ethics Matters*, *supra* note 14, at 72 (“If the autonomous car were most interested in protecting its own occupants, then it would make sense to choose a collision with the lightest object possible.”).

⁷⁹ Goodall, Presentation, *supra* note 50, at 6 (“[A]n automated vehicle programmed to foremost protect the safety of its own passengers can produce morally unacceptable results.”).

⁸⁰ See Michael Anderson & Susan Leigh Anderson, *Robot be Good*, SCI. AM., Oct. 2010, at 72, 77 (“Instilling ethical principles into robots is significant because if people were to suspect that intelligent robots could behave unethically, they could come to reject autonomous robots altogether. The future of AI itself could be at stake.”).

⁸¹ Wadhwa, *supra* note 77; see also Bryant Walker Smith, *Regulation and Risk of Inaction*, in AUTONOMES FAHREN 593, 594 (2015) (“Many vehicle fatalities appear only in local obituaries, but a single automated vehicle fatality would end up on national front pages.”); Tyler Cowen, *Can I See Your License, Registration and C.P.U.?*, N.Y. TIMES (May 28, 2011), <http://www.nytimes.com/2011/05/29/business/economy/29view.html> (“The evening news might show a ‘Terminator’ car spinning out of control and killing a child. There could be demands to shut down the cars until just about every problem is solved. The lives saved by the cars would not be as visible as the lives lost, and therefore the law might thwart or delay what could be a very beneficial innovation.”).

⁸² See *Why Ethics Matters*, *supra* note 14, at 82 (“It’s not outside the realm of possibility to think that the same precautionary backlash won’t happen to the autonomous car industry, if

decision to target the baby carriage and cause one death could mean thousands more accidents and deaths in manually operated vehicle accidents, as well as trillions of dollars in unneeded costs.⁸³ Accordingly, the Shopping Cart Problem introduces the inherent risks in programming crash-optimization algorithms—especially the risk in programming a vehicle to follow a radical self-preservation algorithm.

B. Motorcycle Problem

An autonomous vehicle encounters a situation in which it must strike one of two motorcyclists. To the vehicle's front-left is a motorcyclist who is wearing a helmet. To the vehicle's front-right is a motorcyclist who is not wearing a helmet. Which motorcyclist should the autonomous vehicle strike?⁸⁴

Eighty percent of accidents involving a motorcyclist result in injury or death.⁸⁵ Helmet use reduces the risk of fatality by at least twenty-two percent, and decreases the risk of brain injury by at least forty-one percent.⁸⁶ Although helmet use reduces the risk of serious injury to a motorcyclist in the event of an accident, only nineteen states and the District of Columbia have enacted universal helmet laws that require all motorcyclists to wear a helmet.⁸⁷ Twenty-eight states have enacted partial helmet laws, requiring helmet use by only a portion of the motorcycle-riding population—typically motorcyclists under a certain age.⁸⁸ Even in those states where a helmet is required by state law, motorcyclists only wear their helmets approximately eighty-six percent of the time.⁸⁹ In the states that do not require motorcyclists to wear a helmet, fifty-five percent of motorcyclists wear a helmet anyway.⁹⁰

Based on these statistics, it is clear that the motorcyclist who is not wearing a helmet has a higher chance of injury or death if the

industry doesn't appear to be taking ethics seriously."); Wadhwa, *supra* note 77 ("[The decision to kill the baby] could end autonomous driving in America.").

⁸³ See *supra* Part II.B (discussing the benefits of autonomous vehicles).

⁸⁴ The Motorcycle Problem has been discussed by others. See, e.g., Goodall, Presentation, *supra* note 50, at 2; *The Robot Car of Tomorrow*, *supra* note 6.

⁸⁵ NAT'L HIGHWAY TRAFFIC SAFETY ADMIN., U.S. DEPT OF TRANSP., COUNTERMEASURES THAT WORK: A HIGHWAY SAFETY COUNTERMEASURE GUIDE FOR STATE HIGHWAY SAFETY OFFICES 5-1 (6th ed., 2011), <http://www.nhtsa.gov/staticfiles/nti/pdf/811444.pdf>.

⁸⁶ *Id.* at 5-7.

⁸⁷ *Motorcycles: Motorcycle Helmet Use*, INS. INST. FOR HIGHWAY SAFETY (Jan. 2016), <http://www.iihs.org/iihs/topics/laws/helmetuse/mapmotorcyclehelmets>.

⁸⁸ See *id.*

⁸⁹ NAT'L HIGHWAY TRAFFIC SAFETY ADMIN., *supra* note 85, at 5-7.

⁹⁰ *Id.*

autonomous vehicle hits her.⁹¹ Thus, an autonomous vehicle programmed to reduce the chance of serious injury or death would hit the motorcyclist wearing the helmet.⁹² On the other hand, hitting the motorcyclist who wore a helmet does not seem fair, or even in the best interest of society.⁹³ It seems unfair to punish someone solely because she was responsible, while rewarding someone solely because he was less responsible. Society wants people to wear helmets while riding a motorcycle, and targeting the motorcyclist who wears a helmet may incentivize people not to wear helmets.⁹⁴ Furthermore, in some cases the motorcyclist not wearing the helmet would be breaking the law.⁹⁵ Thus, the Motorcycle Problem introduces another potential issue: an algorithm programmed only to minimize harm may not take into account other important society values, such as fairness.

C. *The Car Problem*

An autonomous vehicle's brakes fail as it approaches a stoplight. Vehicles are crossing the intersection, and there is no way for the autonomous vehicle to safely maneuver without causing an accident. Slightly to the vehicle's left is a 2015 Mercedes-Benz E Class 4-Door sedan with only the driver inside. Slightly to its right is a 2011 Jeep Liberty with only the driver in the car. Which vehicle should the autonomous vehicle strike?⁹⁶

An autonomous vehicle could be programmed to hit the vehicle with the highest safety rating.⁹⁷ The Insurance Institute for

⁹¹ *Why Ethics Matters*, *supra* note 14, at 73 (“[A motorcyclist] without a helmet would probably not survive such a collision.”).

⁹² *See id.*

⁹³ Goodall, *Machine Ethics*, *supra* note 4, at 99 (stating that it is unfair to hit a motorcyclist who wears a helmet instead of an un-helmeted motorist because it would result in less damage); *Why Ethics Matters*, *supra* note 14, at 73.

⁹⁴ *Why Ethics Matters*, *supra* note 14, at 73.

⁹⁵ *See supra* notes 87–89 and accompanying text.

⁹⁶ *The Robot Car of Tomorrow*, *supra* note 6; *see also* White, *supra* note 18 (“[O]ne sticking point is a debate over whether automated driving systems should be programmed to hit a larger sport-utility vehicle or a potentially more vulnerable small car in a case where a crash is unavoidable.”).

⁹⁷ *See* Goodall, *Ethical Decision Making*, *supra* note 1, at 62 (“An automated vehicle could detect nearby vehicle types from its vision system and, in an unavoidable collision, attempt to collide with more compatible vehicles.”); Goodall, *Machine Ethics*, *supra* note 4, at 97 (indicating that a utilitarian autonomous vehicle would strike the vehicle with the highest safety rating); Goodall, Presentation, *supra* note 50, at 2 (“[A]n automated vehicle may consistently choose the SUV with which it has a better crash compatibility . . . rather than the sedan.”); *Why Ethics Matters*, *supra* note 14, at 72 (“[I]t seems to make sense to collide with a safer vehicle . . . over a car not known for crash-safety.”).

Highway Safety (“IIHS”) provides annual highway safety awards based on two aspects of safety: (1) crashworthiness and (2) crash avoidance and mitigation.⁹⁸ Crashworthiness is rated based on five tests—moderate overlap front, small overlap front, side, roof strength, and head restraints—which are used to determine “how well a vehicle protects its occupants in a crash.”⁹⁹ The IIHS then assigns a score of good, acceptable, marginal, or poor based on each of those tests.¹⁰⁰ The crash avoidance and mitigation factor looks at each vehicle’s crash prevention system and the car’s performance in track tests to score each car’s “technology that can prevent a crash or lessen its severity.”¹⁰¹ The IIHS assigns a score of superior, advanced, or basic based on those tests.¹⁰² Each year, the IIHS releases a list of its “Top Safety Pick” and its “Top Safety Pick+” for various categories of vehicles, such as minicars, large family cars, minivans, and large luxury vehicles.¹⁰³ Generally, a Top Safety Pick is a vehicle that received at least “good” ratings in the crashworthiness tests.¹⁰⁴ A Top Safety Pick+ is a Top Safety Pick vehicle with an advanced or superior rating for its front crash prevention.¹⁰⁵

The IIHS scored the 2015 Mercedes E Class four-door Sedan (Mercedes) a 2015 Top Safety Pick+.¹⁰⁶ The Mercedes is equipped with side airbags, front and rear head curtain airbags, front-seat mounted torso airbags, and a driver knee airbag; additionally, it can be equipped with rear seat-mounted torso airbags.¹⁰⁷ The vehicle has a rollover sensor which triggers the side airbags to deploy in the event of a roll over.¹⁰⁸ The Mercedes received the highest crashworthiness scores in all of the tests except in the small overlap front test, where it received an acceptable score for lower leg and foot injury.¹⁰⁹ The

⁹⁸ *IIHS Safety Awards*, INS. INST. FOR HIGHWAY SAFETY, <http://www.iihs.org/iihs/ratings> (last visited Jan. 14, 2016).

⁹⁹ *Id.*

¹⁰⁰ *Id.*

¹⁰¹ *Id.*

¹⁰² *Id.*

¹⁰³ *TOP SAFETY PICKs By Year: 2016*, INS. INST. FOR HIGHWAY SAFETY, <http://www.iihs.org/iihs/ratings/TSP-List> (last visited Jan. 14, 2016) [hereinafter *2016*].

¹⁰⁴ *Id.* However, a car can receive an “acceptable” rating in the small overlap front test and still be a Top Safety Pick. *Id.*

¹⁰⁵ *Id.*

¹⁰⁶ *2015 Mercedes E Class*, INS. INST. FOR HIGHWAY SAFETY, <http://www.iihs.org/iihs/ratings/vehicle/v/mercedes/e-class-4-door-sedan/2015?print-view> (last visited Jan. 14, 2016).

¹⁰⁷ *Id.*

¹⁰⁸ *Id.*

¹⁰⁹ *Id.*

Mercedes received a superior score for the crash avoidance and mitigation factor, but only when it is equipped with the optional front crash prevention system; without that system, the car only scores a basic for this factor.¹¹⁰

The 2011 Jeep Liberty (Jeep) did not receive any top safety pick awards from the IIHS.¹¹¹ The Jeep is equipped with front and head curtain airbags, and it too has a rollover sensor to deploy the side curtain airbags in the event of a rollover.¹¹² It received a good score only for the moderate overlap front and roof strength tests.¹¹³ On the side test, it received a marginal overall score due to the average rating for its structure and safety cage and its poor rating for the risk of driver injury to the torso.¹¹⁴ The notes state that an accident to the side of the Jeep would likely result in rib fractures, internal organ injuries, or both to the driver.¹¹⁵ The Jeep received an average score for the head restraint and seats test.¹¹⁶

Therefore, the Mercedes is equipped with better safety features than the Jeep. But the fact that the Mercedes has better safety features does not necessarily mean that it would be the safest vehicle to hit. The IIHS cautions that “[l]arger, heavier vehicles generally afford more protection than smaller, lighter ones.”¹¹⁷ Thus, a small car that’s a Top Safety Pick+ or Top Safety Pick doesn’t necessarily afford more protection than a bigger car that doesn’t earn the award.”¹¹⁸ The Mercedes is a sedan while the Jeep is a sport utility vehicle—which means that, according to principles of physics, the Jeep would be able to withstand more force than the Mercedes.¹¹⁹

Therefore, if the autonomous vehicle chooses to target the vehicle that has the best safety rating, the autonomous vehicle would hit the Mercedes, even though it may not be able to withstand the impact as well as the Jeep. Conversely, the vehicle could hit the Jeep because it is the larger vehicle—even though the Jeep is not as well equipped

¹¹⁰ *See id.*

¹¹¹ *See TOP SAFETY PICKs By Year: 2011*, INS. INST. FOR HIGHWAY SAFETY, <http://www.iihs.org/iihs/ratings/TSP-List/2011> (last visited Jan. 14, 2016) [hereinafter *2011*]

¹¹² *2011 Jeep Liberty*, INS. INST. FOR HIGHWAY SAFETY, <http://www.iihs.org/iihs/ratings/vehicle/v/jeep/liberty/2011?print-view> (last visited Jan. 14, 2015).

¹¹³ *Id.*

¹¹⁴ *Id.*

¹¹⁵ *Id.*

¹¹⁶ *Id.*

¹¹⁷ *2016*, *see supra* note 103.

¹¹⁸ *Id.*

¹¹⁹ *See The Robot Car of Tomorrow*, *supra* note 6 (“As a matter of physics, [an autonomous vehicle] should choose a collision with a heavier vehicle that can better absorb the impact of a crash.”).

to withstand an accident, especially one to the side of the vehicle. In any event, as in the Motorcycle Problem, it does not necessarily seem fair to target a vehicle because its owner chose a safer car.¹²⁰

An autonomous vehicle that makes its decision based on the safety features of the other vehicles is only considering the amount of harm that could result to the occupants of those vehicles. It is likely that an autonomous vehicle would consider other factors too. One such factor would be the amount of monetary damage that would result to each vehicle if the autonomous vehicle hit it. Depending on which Mercedes E-Class Sedan the person is driving, the manufacturer's suggested retail price (MSRP) is between \$52,650 and \$101,700.¹²¹ The 2011 Jeep Liberty's original MSRP was between \$23,250 and \$28,250, but now this vehicle can be purchased used for around \$11,584 from a dealer.¹²² Therefore, not only would the Jeep better withstand the impact of being hit, the damage to the Jeep would be far less expensive to repair than the damage that would result from hitting the Mercedes.

Another factor an algorithm writer may consider is minimizing the amount of damage to the autonomous vehicle and harm to occupants. Presumably, this factor would lead the autonomous vehicle to always strike the smaller vehicle. As a matter of physics, hitting a smaller vehicle will typically result in less damage to the vehicle and less harm to the occupants.¹²³ Therefore, an autonomous vehicle programmed to minimize its damage and the harm to its occupants would hit the Mercedes.

The autonomous vehicle could also consider various other factors, including the number of people in each vehicle, the seat belt use of the occupants, or perhaps the age of the people in the vehicles.¹²⁴ If the autonomous vehicle considers the demographics of the occupants

¹²⁰ See Goodall, *Machine Ethics*, *supra* note 4, at 97 ("An utilitarian automated vehicle given the choice between colliding with two different vehicles would select the one with the higher safety rating. Although this would maximize overall safety, most would consider it unfair.")

¹²¹ MERCEDES-BENZ, *E-Class Sedan*, <http://www.mbusa.com/mercedes/vehicles/class/class-E/bodystyle-SDN> (last visited Jan. 14, 2016).

¹²² 2011 Jeep Liberty, U.S. NEWS & WORLD REP., http://usnews.rankingsandreviews.com/cars-trucks/Jeep_Liberty/2011/prices/ (last visited Jan. 14, 2016).

¹²³ See *The Robot Car of Tomorrow*, *supra* note 6 ("As a matter of physics, [an autonomous vehicle] should choose a collision with a heavier vehicle that can better absorb the impact of a crash.")

¹²⁴ See ERICA PALMERINI ET AL., REGULATING EMERGING ROBOTIC TECHNOLOGIES IN EUROPE: ROBOTICS FACING LAW AND ETHICS 42 (2014), http://www.robolaw.eu/RoboLaw_files/documents/robolaw_d6.2_guidelinesregulatingrobotics_20140922.pdf ("In complicated cases, making a 'good' decision requires to take into account a broad range of issues."); Goodall, *Ethical Decision Making*, *supra* note 1, at 62 (discussing various factors an autonomous vehicle could consider when determining whether which of two vehicles to strike).

of the other vehicle, it will have a different set of risks to evaluate:

If one driver is a man, and the other a similar-age woman, the woman is 28% more likely to die. If one driver is age 20 and the other age 70, the older driver is three times as likely to die. If one driver is drunk and the other sober, the drunk is twice as likely to die [because alcohol affects many body organs, not just the brain]. If one driver is traveling alone while the other has a passenger, the lone driver is 14% more likely to die than the accompanied driver, because the accompanied driver is in a vehicle heavier by the mass of its passenger.¹²⁵

An autonomous vehicle with face recognition technology could detect these characteristics about the occupants of the other vehicles.¹²⁶ As discussed in Part IV, these factors would be relevant to a consequentialist autonomous vehicle concerned with minimizing harm. But most people would find an autonomous vehicle that targets people on these grounds disturbing and unethical,¹²⁷ and the ethical standards for engineers would prohibit them from programming their vehicles to make collision decisions based on such grounds.¹²⁸

D. The Tunnel Problem

An autonomous vehicle is travelling along a single lane mountain road that is fast approaching a narrow tunnel. Just before the car enters 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 child, or swerve into the wall on either side of the tunnel, thus killing its operator. How should the car react?

Jason Millar created the Tunnel Problem in an article written for

¹²⁵ Goodall, *Ethical Decision Making*, *supra* note 1, at 62 (alterations in original).

¹²⁶ *Id.* (“An automated vehicle that used facial recognition technology could estimate the gender, age, or level of inebriation of nearby drivers and passengers, and adjust its objective function accordingly.”).

¹²⁷ *See id.* at 62 (noting that many would find an autonomous vehicle using these factors disturbing); *The Robot Car of Tomorrow*, *supra* note 6 (stating that decisions on similar grounds would be troubling).

¹²⁸ The Institute of Electrical and Electronics Engineers (IEEE) requires its members “to treat fairly all persons and to not engage in acts of discrimination based on race, religion, gender, disability, age, national origin, sexual orientation, gender identity, or gender expression.” INST. OF ELECTRICAL & ELECTRONICS ENGINEERS, 7.8 IEEE CODE OF ETHICS, <http://www.ieee.org/about/corporate/governance/p7-8.html> (last visited Jan. 14, 2016).

Robohub.¹²⁹ Thereafter, the Open Roboethics Initiative polled Robohub's readers about how they would make the decision in the Tunnel Problem.¹³⁰ Of those polled, sixty-four percent stated that the car should kill the child;¹³¹ forty-eight percent said the decision was easy, and only twenty-four percent thought the decision was difficult.¹³² The Open Roboethics Initiative also asked for the reasoning behind each person's decision.¹³³ Of those who said they would kill the child, thirty-two percent based that decision on the child's level of fault for being in the road, thirty-three percent believed the car should be biased in favor of its passengers, and thirteen percent responded that they would always choose themselves.¹³⁴ Of those who said save the child and kill themselves, thirty-six percent provided altruistic reasons, fourteen percent based their decision on the child having more time to live, and eleven percent said they did not want to live with the guilt of killing a child.¹³⁵

The Tunnel Problem presents the unique situation where the autonomous vehicle must decide whether to kill the operator or a third party. If the autonomous vehicle chooses to kill its owner, it would mean that a company has programmed a vehicle to kill its customers—at least in certain situations. From an economic standpoint, car manufacturers probably could not make the decision to kill its consumers to save one person—or any number of people, for that matter. If a consumer knew that, for example, ACME Car Manufacturing Company made the decision to program its vehicle to sacrifice the occupant in order to save any one person, and XYZ Car Company was willing to save the occupants over third parties, then it seems likely that people would purchase their autonomous vehicles from XYZ.¹³⁶ In addition, XYZ would probably run advertisements

¹²⁹ Millar, *supra* note 15.

¹³⁰ Open Roboethics Initiative, *If Death by Autonomous Car is Unavoidable, Who Should Die? Reader Poll Results*, ROBOHUB (June 23, 2014), <http://robohub.org/if-a-death-by-an-autonomous-car-is-unavoidable-who-should-die-results-from-our-reader-poll/> [hereinafter *Who Should Die?*].

¹³¹ *Id.*

¹³² *Id.*

¹³³ Open Roboethics Initiative, *My (Autonomous) Car, My Safety: Results From Our Reader Poll*, ROBOHUB (June 30, 2014), <http://robohub.org/my-autonomous-car-my-safety-results-from-our-reader-poll/> [hereinafter *My (Autonomous) Car*].

¹³⁴ *Id.*

¹³⁵ *Id.*

¹³⁶ See Goodall, Presentation, *supra* note 50, at 6 ("Understandably, the owner of an automated vehicle may have a strong incentive to maximize the safety of his own vehicle and its occupants over the safety of other roadway users.").

pointing out that distinction: “ACME will kill your family” or “At XYZ, we care about *you*.” Thus, it seems likely that the market will force car manufacturers to be biased in favor of their consumers.¹³⁷

The Tunnel Problem also presents a recurring theme in the analysis of crash-optimization algorithms: what impact should “fault” on the part of a third party have on the autonomous vehicle’s decision? After all, the reason the autonomous vehicle has to make this “who to kill” decision is because the child ran into the street. For those uncomfortable with the thought of an autonomous vehicle potentially killing a child, the Tunnel Problem could be modified to replace the child with a psychotic adult who believes that all autonomous vehicle owners should die. In this modified version, the fact that someone is trying to get the vehicle to destroy itself and kill its occupants could change people’s opinion about which decision the autonomous vehicle should make. In any event, there is still the technical question of whether autonomous vehicles could be programmed to take fault into account—and if they could, whether society would find this desirable.

E. The Bridge Problem

An autonomous vehicle with only one occupant is traveling along a narrow, two-lane bridge. A school bus full of children is travelling in the opposite direction. The driver of the school bus is drowsy. As the school bus and the autonomous vehicle approach each other, the driver of the school bus starts to doze off, causing the school bus to drift into the autonomous vehicle’s lane. The autonomous vehicle has two options: (1) crash into the school bus; or (2) drive off the bridge, killing its occupant.¹³⁸

The Bridge Problem raises the issue of self-sacrifice more forcefully

¹³⁷ This raises a question as to whether the manufacturer should make the decision. See Millar, *supra* note 15 (arguing that the manufacturer should not make these life or death decisions); *infra* Part VI.

¹³⁸ The Bridge Problem was introduced by Gary Marcus in an article for the *New Yorker*. Gary Marcus, *Moral Machines*, NEW YORKER (Nov. 24, 2012), <http://www.newyorker.com/news/news-desk/moral-machines> (“Your car is speeding along a bridge at fifty miles per hour when errant [sic] school bus carrying forty innocent children crosses its path. Should your car swerve, possibly risking the life of its owner (you), in order to save the children, or keep going, putting all forty kids at risk?”); see also Goodall, *Ethical Decision Making*, *supra* note 1, at 60; Patrick Lin, *The Ethics of Saving Lives with Autonomous Cars is Far Murkier Than You Think*, WIRED (July 30, 2013), <http://www.wired.com/2013/07/the-surprising-ethics-of-robot-cars/> [hereinafter *The Ethics of Saving Lives*] (discussing the Bridge Problem); Clive Thompson, *Relying on Algorithms and Bots Can Be Really, Really Dangerous*, WIRED (Mar. 25, 2013), <http://www.wired.com/2013/03/clive-thompson-2104/> (discussing Marcus’s hypothetical).

than the Tunnel Problem. An autonomous vehicle that cannot commit self-sacrifice may end up killing all of the children and the bus driver, while also damaging itself, and harming or potentially killing its occupant.¹³⁹ Nothing good will result from a head-on collision with a full school bus on a narrow, two-lane bridge. Thus, self-sacrifice by the autonomous vehicle will likely result in less harm. Although the autonomous vehicle will be destroyed and its occupant killed, it is likely that no one else will be harmed and no other damage will occur. Therefore, an autonomous vehicle programmed to minimize harm would probably commit self-sacrifice.¹⁴⁰

In this example, the autonomous vehicle is again “innocent” in bringing about the harm. The school bus driver is at fault for whatever harm occurs. But this type of “fault” is different than the type of “fault” in the Tunnel Problem: none of the children are at fault for the bus driver’s drowsiness, and they are also innocent. Thus, fault is a complicated concept, and may not necessarily be a relevant factor.

What should also be noted about this Problem is that the preferred decision is also fact-specific. An autonomous vehicle should not be programmed to avoid accidents with all school buses: if the school bus’s only occupant is the bus driver, it is likely that self-sacrifice would not be the best decision.¹⁴¹ If the autonomous vehicle drives off the bridge, one death is guaranteed, but if it collides with the school bus, perhaps only serious injury will result.¹⁴²

F. The Trolley Problem

An operator is driving her autonomous car in manual mode and is in control of the vehicle. Either intentionally or not—she could be homicidal or simply inattentive—she is about to run over and kill five pedestrians. Her car’s crash-avoidance system detects the possible accident and activates, forcibly taking control of the car. To avoid this disaster, the car swerves in the only direction it can—say, to the right. But on the car’s right is a single pedestrian who the car strikes

¹³⁹ It should be noted, however, that the autonomous vehicle could drive off the bridge, realize it is doing so, and roll down the windows to give its operator the best possible chance to survive.

¹⁴⁰ See *Why Ethics Matters*, *supra* note 14, at 77 (stating that human drivers might sacrifice themselves to save a school bus full of children).

¹⁴¹ See *id.*

¹⁴² See *id.*

and kills.¹⁴³

This hypothetical is named after the famous philosophical thought-experiment created by Philippa Foot.¹⁴⁴ In the traditional formulation of this Problem, the trolley conductor must decide whether to stay on the same track and kill five people or steer her runaway train onto another track where one person will be killed.¹⁴⁵ Judith Jarvis Thomson created the bystander variation, where the trolley is again faced with either killing five persons by staying on its track or killing one person by switching tracks.¹⁴⁶ In this variation, however, the trolley driver has passed out and his brakes have failed.¹⁴⁷ A bystander is walking along the track near the switch that allows her to change the tracks to kill the one person.¹⁴⁸ The bystander variation differs from the original trolley problem in two main respects.¹⁴⁹ First, the trolley driver is in a special position to protect those who may be harmed by the trolley.¹⁵⁰ Second, if the trolley driver decides to drive the trolley into the five people, he would actively kill them by running into them with his vehicle; on the other hand, if the bystander decides not to flip the switch, the five people would die due to the bystander's inaction.¹⁵¹ The Foot Trolley Problem would be more reminiscent of a situation where the brakes fail on a traditional vehicle and the driver must determine whether to turn the steering wheel. The bystander variation is more analogous to the choice confronting the autonomous vehicle manufacturer when deciding how to program its vehicles.

What makes the Trolley Problem potentially troublesome is not the fact that the autonomous technology took control of the vehicle to save the human driver from causing an accident or killing people. Society will want the autonomous technology to intervene when it can safely avert an accident, and automobiles are already equipped with technology to protect people when they make mistakes or fall asleep while driving.¹⁵² For example, Volvo has implemented new

¹⁴³ The autonomous vehicle variation of the trolley problem was created by Patrick Lin. *Id.* at 79.

¹⁴⁴ Philippa Foot, *The Problem of Abortion and the Doctrine of the Double Effect*, 5 OXFORD REV. 5, 8 (1967).

¹⁴⁵ *Id.*

¹⁴⁶ Judith Jarvis Thomson, Comment, *The Trolley Problem*, 94 YALE L.J. 1395, 1397 (1985).

¹⁴⁷ *Id.*

¹⁴⁸ *Id.*

¹⁴⁹ *Id.*

¹⁵⁰ *Id.*

¹⁵¹ *Id.*

¹⁵² See, e.g., Bilger, *supra* note 16, at 107 (describing Volvo's anti-accident prevention system).

safety systems to address the four most common causes of accidents: distraction, drowsiness, drunkenness, and driver error.¹⁵³ The accident prevention system includes sensors to alert the driver if the car crosses a lane line without a blinker; furthermore, if the driver's behavioral patterns indicate drowsiness, the technology suggests that the driver take a break.¹⁵⁴ If those systems fail, the car tightens seat belts and will attempt to stop the car prior to a collision.¹⁵⁵ Volvo hopes that its safety features will eventually be so robust that no one will be seriously injured or killed in one of its vehicles.¹⁵⁶ Therefore, it is clear that technology's ability to prevent an accident should be fostered by society.

However, what is potentially troublesome with the Trolley Problem is that the autonomous technology killed someone by preventing the death of five persons. Decisions of life or death are inherently moral ones, and for some people, the thought of a "robot car" making those decisions seems repulsive.¹⁵⁷ They would suggest that deeply ethical decisions should be made by the operator, not the programmer.¹⁵⁸ The ethical basis of the Trolley Problem will be discussed in Part IV, and Part VI will discuss who should make these decisions.

The Trolley Problem also introduces an important legal distinction between action and inaction. As discussed more fully in Part V, the law does not typically require people to act.¹⁵⁹ Thus, in the bystander problem, the bystander who walks by the switch without flipping it is not legally responsible for the deaths of the five people who are killed. Likewise, if the autonomous technology does not take control of the vehicle to save the five pedestrians, the manufacturer is not legally responsible for their deaths.¹⁶⁰ Instead, the human driver will be criminally and civilly responsible.¹⁶¹ On the other hand, if the autonomous technology takes control of the vehicle to save five lives and ends up killing one person, the manufacturer is civilly and

¹⁵³ *Id.*

¹⁵⁴ *Id.*

¹⁵⁵ *Id.*

¹⁵⁶ Matt McFarland, *Peak Fender-Bender: Technology Can Prevent Car Crashes, if Consumers Will Buy in*, WASH. POST (Feb. 10, 2015), <http://www.washingtonpost.com/blogs/innovations/wp/2015/02/10/peak-fender-bender-technology-can-prevent-car-crashes-if-consumers-will-buy-in/>.

¹⁵⁷ See Jason Millar, *You Should Have a Say in Your Robot Car's Code of Ethics*, WIRED (Sept. 2, 2014), <http://www.wired.com/2014/09/set-the-ethics-robot-car/>.

¹⁵⁸ *See id.*

¹⁵⁹ *See infra* Part V.A.

¹⁶⁰ *See Why Ethics Matters*, *supra* note 14, at 79 ("If the car does not wrestle control from the human driver, then it (and the OEM) would perhaps not be responsible for the deaths of the five pedestrians while you were driving the car; it is merely letting those victims die.").

¹⁶¹ *See id.*

perhaps even criminally responsible for the death of the person killed.¹⁶² Therefore, the law creates an enormous incentive for the automaker to program its autonomous technology not to intervene in the Trolley Problem—or for that matter, in any case where an accident would result by avoiding a greater harm caused by the human driver.

IV. MORAL PHILOSOPHY

Machine ethics, robot ethics, and machine morality are all terms used to describe a nascent field of ethics concerned with “the practical challenge of building (ro)bots which explicitly engage in making moral decisions.”¹⁶³ Machine ethics distinguishes between top-down and bottom-up approaches,¹⁶⁴ but many experts believe ethical robots will require a combination of both approaches.¹⁶⁵ A top-down approach implements normative ethical theories into robotics and requires robots to follow those rules.¹⁶⁶ Examples of top-down rules include Asimov’s laws, utilitarianism, Kantianism, egoism, virtue ethics, and the Ten Commandments.¹⁶⁷ Conversely, “[i]n bottom-up approaches to machine morality, the emphasis is placed on creating an environment where an agent explores courses of action and learns and is rewarded for behavior that is morally praiseworthy.”¹⁶⁸ Child development is an example of a bottom-up approach to ethics.¹⁶⁹ Thus, the robot discovers and constructs its ethical principles.¹⁷⁰

This Part utilizes a top-down approach and examines the feasibility of programming a crash-optimization algorithm to follow normative ethics, which describes how people ought to act, not how they actually act.¹⁷¹ Normative ethical theories have been debated

¹⁶² See *id.* (“[I]f the car does take control and make a decision that results in the death of a person, then it (and the OEM) becomes responsible for killing a person.”).

¹⁶³ Wallach, *The Challenge of Moral Machines*, *supra* note 3, at 6.

¹⁶⁴ Wendell Wallach, *Robot Minds and Human Ethics: The Need for a Comprehensive Model of Moral Decision Making*, 12 ETHICS & INFO. TECH. 243, 247 (2010) [hereinafter Wallach, *Robot Minds and Human Ethics*].

¹⁶⁵ See, e.g., WALLACH & ALLEN, *supra* note 12, at 117 (noting that ethical robots will require top-down and bottom-up approaches to ethics).

¹⁶⁶ See Wallach, *Robot Minds and Human Ethics*, *supra* note 164, at 247.

¹⁶⁷ WALLACH & ALLEN, *supra* note 12, at 84 (“Top-down ethical systems might come from a variety of sources, including religion, philosophy, and literature. Examples include the Golden Rule, the Ten Commandments, consequentialist or utilitarian ethics, Kant’s moral imperative, legal and professional codes, and Asimov’s Three Laws of Robotics.”).

¹⁶⁸ *Id.* at 80.

¹⁶⁹ *Id.*

¹⁷⁰ See *id.*

¹⁷¹ Goodall, Presentation, *supra* note 50, at 4.

for millennia, and mankind has yet to come to a consensus on which theory is “right.”¹⁷² Plato and Aristotle believed that ethics depended on the character of the ethical actor—commonly referred to as “virtue ethics.”¹⁷³ A virtue ethicist tries to produce an “ideal person” by determining the characteristics that make a person virtuous; once those characteristics are outlined, the moral agent “aspire[s] to be an ideal person” by acquiring those characteristics.¹⁷⁴ As such, virtue ethics combines a top-down approach—the characteristics—with a bottom-up approach—the acquisition of those characteristics—to robot ethics.¹⁷⁵ This combination has led some roboethicists to believe that virtue ethics is a promising ethical system for robots.¹⁷⁶ However, character traits, such as courage, patience, optimism, and industry, provide little insight into how an autonomous vehicle’s crash-optimization algorithm should be programmed.

In light of the shortcomings to programming virtue ethics into an autonomous vehicle, a programmer would need to consider other ethical theories, one such theory is the “Divine Command Theory.” Under this theory, God determines what actions are morally right or wrong.¹⁷⁷ For example, in the Judeo-Christian tradition, God established the rule “Thou shalt not kill.”¹⁷⁸ Thus, an autonomous vehicle would be programmed not to kill. However, this seemingly bright-line rule is not as clear as it first seems. First, the prohibition against killing another person does not necessarily apply to all killings.¹⁷⁹ Second, this rule provides no method to resolve a situation when someone must be killed, such as in the Tunnel Problem.¹⁸⁰ Given these complications, the programming of a religious autonomous vehicle would require consultation with religious experts; this raises the additional question of what religious expert

¹⁷² See, e.g., Anderson & Anderson, *supra* note 80, at 75 (“Assuming that it is possible to give ethical rules to robots, whose ethical rules should those be? After all, no one has yet been able to put forward a general set of ethical principles for real-live humans that is accepted universally.”).

¹⁷³ See LOUIS P. POJMAN, *ETHICS: DISCOVERING RIGHT AND WRONG* 156 (Steve Wainwright et al. eds., 5th ed. 2006).

¹⁷⁴ See *id.* at 160.

¹⁷⁵ See, e.g., WALLACH & ALLEN, *supra* note 12, at 117–24 (discussing virtue ethics); Abney, *supra* note 3, at 51 (“[V]irtue ethics . . . [i]s the best approach to robot ethics.”).

¹⁷⁶ See, e.g., WALLACH & ALLEN, *supra* note 12, at 118; Abney, *supra* note 3, at 51.

¹⁷⁷ JAMES RACHELS & STUART RACHELS, *THE ELEMENTS OF MORAL PHILOSOPHY* 50 (Michael Ryan et al. eds., 6th ed. 2010).

¹⁷⁸ See *Exodus* 20:13 (King James).

¹⁷⁹ The Bible provides ample examples of killing that were not proscribed by the Commandment against killing. See, e.g., *Exodus* 22:24 (King James).

¹⁸⁰ The Bible proscribes killing and suicide. See, e.g., *Exodus* 20:13 (King James); 1 *Corinthians* 3:17 (King James).

to consult. In the Christian faith alone, countless denominations exist.¹⁸¹ Because of the multitude of questions that must be answered before programming a religious autonomous vehicle, the Divine Command Theory seems impractical to use for crash-optimization algorithms.¹⁸²

A third ethical theory that could be programmed into the crash-optimization algorithm is “ethical egoism.” Under this theory, morality requires each person to always promote her self-interest; thus, each person just asks herself what she feels like doing and that decision is moral.¹⁸³ To program an ethical egoist autonomous vehicle, car manufacturers would need to let every individual car owner choose her own moral decisions so that she could decide what is in her best interest. As discussed in Part VI, car manufacturers or the government would foreclose car owners from some ethical criteria: no one will be able to program a vehicle to target someone because of her race, gender, religion, or other arbitrary ground.¹⁸⁴ In addition to some choices being foreclosed, this theory is impractical; it would be difficult, if not impossible, for a car manufacturer to program each vehicle to follow that car owner’s self-interest. Because of these shortcomings, ethical egoism is not a realistic choice for programming autonomous vehicles.

Two remaining broad ethical theories could be programmed into autonomous vehicles: utilitarianism (both act and rule utilitarianism) and Deontology (Kantianism). These two theories are generally considered “rivals” and are useful to show the intricate ethical issues facing the programming of an ethical crash-optimization algorithm.¹⁸⁵ The remaining discussion in this Part explores those theories, as well as their interplay with autonomous vehicles and the ethical dilemmas outlined in Part III.¹⁸⁶

¹⁸¹ See, e.g., *Grouping Christian Denominations into Families*, ONTARIO CONSULTANTS ON RELIGIOUS TOLERANCE, http://www.religioustolerance.org/chr_deno.htm (last updated Nov. 6, 2006).

¹⁸² For interesting articles discussing Divine Command Theory and robot ethics, see Selmer Bringsjord & Joshua Taylor, *The Divine-Command Approach to Robot Ethics*, in *ROBOT ETHICS: THE ETHICAL AND SOCIAL IMPLICATIONS OF ROBOTICS* 85, 85 (Patrick Lin et al. eds., 2012); James Hughes, *Compassionate AI and Selfless Robots: A Buddhist Approach*, in *ROBOT ETHICS: THE ETHICAL AND SOCIAL IMPLICATIONS OF ROBOTICS* 69, 78–80 (Patrick Lin et al. eds., 2012).

¹⁸³ FRED FELDMAN, *INTRODUCTORY ETHICS* 80 (1978); POJMAN, *supra* note 173, at 81; RACHELS & RACHELS, *supra* note 177, at 63.

¹⁸⁴ See *infra* Sections VI.B.2, VI.B.3.

¹⁸⁵ See, e.g., WALLACH & ALLEN, *supra* note 12, at 85 (“[Utilitarianism and deontology] are two ‘big picture’ rivals for what th[e] general [ethical] principle[s] should be.”).

¹⁸⁶ Part IV does not address the shortcomings or objections to the respective theories, except for when a particular objection is relevant to autonomous vehicles.

A. Utilitarianism

Utilitarianism is a consequentialist or teleological moral theory,¹⁸⁷ and it has two main features: a consequentialist principal and a utility principle.¹⁸⁸ The consequentialist principal states that the rightness of an action is based on the consequences of that action; as such, utilitarianism is concerned with the ends of an action, not the means.¹⁸⁹ The utility principal focuses on the type of state—for example, pleasure or happiness—that is used to measure the consequences.¹⁹⁰

There are two principal utilitarian theories: (1) act utilitarianism; and (2) rule utilitarianism.¹⁹¹ Under act utilitarianism, “[a]n act is right if and only if it produces the greatest happiness of the greatest number.”¹⁹² Conversely, under rule utilitarianism, “[a]n act is right if and only if it is required by a rule that is itself a member of a set of rules whose acceptance would lead to greater utility for society than any available alternative.”¹⁹³ Therefore, act and rule utilitarianism are similar in that they focus on maximizing utility, but they differ because the former focuses solely on maximizing utility of each individual act while the latter is concerned with maximizing utility of repeatedly performed acts.¹⁹⁴

Because act utilitarianism focuses on each individual act, every decision involves a moral “calculation” of the utility gained or lost by such an action.¹⁹⁵ Seeing as how autonomous vehicles will be operated by a computer, the use of calculations is appealing—at least at first.¹⁹⁶ A computer can aggregate utility more quickly and

¹⁸⁷ See WALLACH & ALLEN, *supra* note 12, at 85.

¹⁸⁸ POJMAN, *supra* note 173, at 107.

¹⁸⁹ *Id.* at 107–08.

¹⁹⁰ *Id.* at 108 (“The utility, or *hedonist*, principle states that the only thing that is good in itself is some specific type of state (e.g., pleasure, happiness, welfare).”). Jeremy Bentham, a father of the utilitarian school, stated that society has two sovereign masters: “Nature has placed mankind under the governance of two sovereign masters, *pain* and *pleasure*. It is for them alone to point out what we ought to do, as well as to determine what we shall do.” JEREMY BENTHAM, AN INTRODUCTION TO THE PRINCIPLES OF MORALS AND LEGISLATION 1 (Hafner Publ’g, Co., 1948) (1823). John Stuart Mill, another father of the utilitarian school, focused on happiness, which he defined as the “pleasure, and the absence of pain.” JOHN STUART MILL, UTILITARIANISM 16–17 (Prometheus Books 1987) (1863).

¹⁹¹ POJMAN, *supra* note 173, at 110.

¹⁹² FELDMAN, *supra* note 183, at 27.

¹⁹³ POJMAN, *supra* note 173, at 111.

¹⁹⁴ See FELDMAN, *supra* note 183, at 61–62.

¹⁹⁵ See POJMAN, *supra* note 173, at 110.

¹⁹⁶ See WALLACH & ALLEN, *supra* note 12, at 86; James Gips, *Towards the Ethical Robot*, in ANDROID EPISTEMOLOGY 243, 248–49 (Kenneth M. Ford et al. eds., 1995) (“At first glance, consequentialist theories might seem the most ‘scientific,’ the most amenable to

accurately than a human.¹⁹⁷ Unlike the human who lacks the time and the ability to aggregate happiness on every decision she makes while driving,¹⁹⁸ an autonomous vehicle could constantly run calculations to ensure that each of its decisions maximizes utility. To address the moral dilemmas from Part III, the autonomous vehicle would need to be able to correctly identify that it has encountered a situation in which its crash-optimization algorithm is necessary.¹⁹⁹ Once it correctly identifies the situation, the car would then generate the possible alternative choices available to it.²⁰⁰ The vehicle would then determine the results of each choice.²⁰¹ And finally, the autonomous vehicle would aggregate the utility of each result to determine which choice maximizes utility.²⁰²

Utilitarianism can be applied to the moral dilemmas from Part III. In the Shopping Cart Problem, the autonomous vehicle would calculate the expected amount of pain that would result from each of its options: colliding into the grocery store, the shopping cart, or the baby carriage. A collision with the grocery store would likely cause property damage to the building, shoppers, the car itself, and its occupant. If the autonomous vehicle hit the shopping cart, the car would likely destroy the groceries and the shopping cart, as well as cause minimal damage to the autonomous vehicle and minimal harm to the occupant. The collision with the baby carriage would likely result in the death of the baby, destruction of the baby carriage, and de minimis damage to the autonomous vehicles and little, if any, harm to vehicle's occupant. As is evident, the utilitarian autonomous vehicle would likely strike the shopping cart because it would result in the least amount of harm.

In the Motorcycle Problem, the autonomous vehicle would calculate

implementation in a robot.”); Goodall, *Machine Ethics*, *supra* note 4, at 99 (“The advantage of [utilitarianism] is that it is easily computable.”); Goodall, Presentation, *supra* note 50, at 4 (“[Utilitarianism] is popular among computer scientists due to its computability . . .”).

¹⁹⁷ See WALLACH & ALLEN, *supra* note 12, at 87; Gips, *supra* note 196, at 248–49 (“At first glance, consequentialist theories might seem the most ‘scientific,’ the most amenable to implementation in a robot.”); Goodall, *Machine Ethics*, *supra* note 4, at 99 (“The advantage of [utilitarianism] is that it is easily computable.”); Goodall, Presentation, *supra* note 50, at 4 (“[Utilitarianism] is popular among computer scientists due to its computability . . .”).

¹⁹⁸ Goodall, *Ethical Decision Making*, *supra* note 1, at 60 (“Thus [autonomous vehicles] will be able to overcome many of the limitations experienced by humans. If a crash is unavoidable, a computer can quickly calculate the best way to crash on the basis of a combination of safety, the likelihood of the outcome, and certainty in measurements much faster and with greater precision than a human can.”).

¹⁹⁹ *Id.* at 61.

²⁰⁰ *Id.*

²⁰¹ *Id.*

²⁰² See *id.*

the amount of harm that would result from hitting the motorcyclist with a helmet in comparison to striking the motorcyclist without a helmet. As indicated in the introduction of the Motorcycle Problem, helmet use greatly reduces the risk of death and brain injury that results from an accident.²⁰³ Therefore, less harm would likely result if the autonomous vehicle struck the motorcyclist who wore a helmet. As such, the act-utilitarian autonomous vehicle would hit the helmeted motorcyclist.²⁰⁴ Whether that decision is fair does not matter to the cold calculation of act utilitarianism. A rule utilitarian, however, would argue that helmet use actually increases happiness; therefore, the rule should be one that promotes helmet use. A rule that autonomous vehicles should always strike a helmeted motorcyclist over an unhelmeted motorcyclist would lead to less happiness: it would incentivize motorcyclists not to wear helmets because an autonomous vehicle would be programmed to avoid striking them. Conversely, a rule that protects helmet wearers may lead people to wear helmets. Thus, the rule utilitarian would argue that, although in some instances targeting the helmet wearer would maximize happiness, a rule protecting motorcyclists who wore helmets would actually promote the greatest happiness.

In the Car Problem, an act-utilitarian autonomous vehicle would calculate the amount of damage and harm that would result from hitting each vehicle. The act-utilitarian car would take into account the safety ratings, size, and value of each vehicle, as well as other factors such as the number of people in each vehicle, and the seat belt use of the occupants. In addition, the car could consider the age and gender of the occupants, which could be indicative of the ability of an occupant to withstand the accident.²⁰⁵ The autonomous vehicle would weigh each factor and determine which vehicle to hit. Similar to the rule utilitarian's objections in the Motorcycle Problem, it is likely that a rule utilitarian would argue that the autonomous should not target the vehicle with the highest safety rating or the vehicle with its occupants wearing seat belts—those decisions would discourage happiness-promoting behavior and result in less happiness.

²⁰³ See *supra* notes 87, 92–93 and accompanying text.

²⁰⁴ See *supra* notes 87, 92–93 and accompanying text.

²⁰⁵ But note that a basic tenet of utilitarianism is that each person is valued the same, and thus, at least theoretically, the age of the occupants would not matter. See RACHELS & RACHELS, *supra* note 177, at 109 (stating that everyone is valued the same). Of course, people do not treat everyone the same. A person would certainly value family members and close friends more than a stranger. From a utilitarian standpoint, it seems clear that a world class physician would provide more utility than someone who is suffering from a terminal disease.

In the Tunnel Problem, the autonomous vehicle would need to determine who it should kill: the child or the occupant.²⁰⁶ The only material difference between the two is that killing the occupant will result in the total destruction of the autonomous vehicle, while killing the child will result in some, but not total, destruction of the vehicle. Therefore, it seems likely that less pain would occur from hitting the child. A rule utilitarian would likely agree with the result, but for a different reason: a rule that killed the operator would lead to less people purchasing autonomous vehicles. Very few people will likely want an autonomous vehicle that was programmed to sacrifice themselves or their family—especially if they were sacrificed to save one person. Because of the projected benefits of autonomous vehicles discussed in Part II, the rule utilitarian would likely advocate for a rule that leads people to purchase these cars. In this case, that rule would be to program the autonomous vehicle with a bias in favor of the operator.

In the Bridge Problem, if the autonomous vehicle is programmed such that it does not avoid the school bus, the decision will likely result in the death of all the children and the driver of the school bus; damage to or destruction of the school bus; serious damage to the autonomous vehicle; and harm to the operator.²⁰⁷ Instead of staying the course, the autonomous vehicle could drive off the bridge, killing the occupant and destroying the autonomous vehicle; in such a situation, no one on the school bus would be harmed and the bus would not be damaged. Because hitting the bus would cause more harm, the act-utilitarian autonomous vehicle would commit self-sacrifice and drive off the bridge. A rule utilitarian would likely agree with the act-utilitarian. An ultra-selfish autonomous vehicle that would not sacrifice itself when the risk of death and harm to its occupant was so high would likely create much unhappiness. And although many people may not want to purchase autonomous vehicles that are willing to sacrifice themselves when the choice was one person or the occupant, it is unlikely that people would have the same reservations as the number of persons killed increases—indeed, society may not even want people to program a vehicle that is that selfish.

In the Trolley Problem, the programmer has the choice of whether the autonomous vehicle should prevent five deaths when doing so will

²⁰⁶ See *supra* Section III.D.

²⁰⁷ See *supra* Section III.E.

result in one death.²⁰⁸ This is a simple choice for a utilitarian autonomous vehicle: it would take control of the vehicle and kill the one person—five lives are better than one life.

Although utilitarianism can be applied to the moral dilemmas from Part III, a utilitarian autonomous vehicle creates unique issues that warrant further discussion. Such vehicles suffer from a variation of the frame problem.²⁰⁹ The frame problem is “[t]he problem of determining, from all the information that can in principle bear, what is relevant to the cognitive task at hand.”²¹⁰ And this problem invites a computational frame problem, which is how a “cognitive system tractably delimits (i.e., frames) what gets considered in a given cognitive task.”²¹¹ The frame problem of crash-optimization algorithms is what factors an autonomous vehicle should consider when determining what decision to make and how to program those factors so that the car applies them when needed.²¹² For example, some facts that the vehicle needs to know to operate are irrelevant for determining which choice to make in these situations: the autonomous vehicle does not need to know that to travel “home” it would need to take X, Y, and Z roads; that it could drive from New York City, New York to Los Angeles, California in approximately forty hours; or that a red light means stop. Although all these facts are correct, they are irrelevant information in any of the moral dilemmas. Humans are capable of framing relevant information when faced with a moral dilemma, or more broadly any decision making that a human undertakes.²¹³ A computer may not have the same framing ability, and without some ability to frame the proper criteria, the autonomous vehicle may be unable to determine what course of action to take in time to actually implement its decision.²¹⁴

²⁰⁸ See *supra* Section III.F.

²⁰⁹ See Abney, *supra* note 3, at 45.

²¹⁰ Sheldon J. Chow, *What's the Problem with the Frame Problem?*, 4 REV. PHIL. & PSYCH. 309, 312 (2013).

²¹¹ *Id.* at 314.

²¹² See *Why Ethics Matters*, *supra* note 14, at 74–75 (discussing the various factors at issue when determining whether to hit a deer).

²¹³ See Chow, *supra* note 210, at 312–13 (“[H]umans have an astonishing ability for determining (with reasonable levels of success) what is relevant in much of their reasoning, and fixating thereon, without having to spend much time deciding what is and is not relevant, or wasting time cogitating on irrelevant details.”); Wallach, *The Challenge of Moral Machines*, *supra* note 3, at 7 (“[F]or most moral challenges that humans respond to in daily life, the capacity to reason is dependent on a vast reservoir of knowledge and experience.”).

²¹⁴ See Daniel Dennett, *Cognitive Wheels: The Frame Problem of AI*, in MINDS, MACHINES AND EVOLUTION 129, 150 (Christopher Hookway ed., 1984). In that article, Dennett tells a story involving a robot that is supposed to retrieve a spare battery from a room; in that room is also a bomb that was set to detonate. *Id.* at 129. The story discusses the problem that

In addition to suffering from a variation of the frame problem, an act-utilitarian autonomous vehicle may intentionally violate traffic laws when doing so will lead to greater happiness—this would likely be a problem for the vehicles.²¹⁵ For example, if an act-utilitarian autonomous vehicle is traveling on a long, flat road in the middle of nowhere and no vehicles are around, why would the autonomous vehicle drive the speed limit of forty-five miles per hour? Or if an autonomous vehicle is stuck behind a slow driver of a traditional vehicle in a “no passing zone,” and there is no oncoming traffic and the vehicle determines that it can safely pass the vehicle, why would it follow the law? If breaking the law led to more happiness in those two examples, the act-utilitarian autonomous vehicle would break the law. However, it is unlikely that the government would allow a car to violate the law: states that have passed autonomous vehicle laws require the vehicles to follow traffic laws.²¹⁶ A rule utilitarian would likely agree with the government’s position because society would not have enacted traffic laws that do not increase happiness. If society did enact such a law, the proper course of action is repealing or amending that law. A rule utilitarian would also assert that if general law breaking became common practice and went unpunished, such rule-breaking would invite contempt for the law and others to engage in law-breaking, too. Therefore, a rule-utilitarian autonomous vehicle would likely follow traffic laws.

A third problem is that, depending on the value assigned to a human life, an act-utilitarian autonomous vehicle may elect to kill a person to protect property. The Environmental Protection Agency, the Food and Drug Administration, and the Transportation Department value one human life at \$9.1, \$7.9, and \$6.1 million,

programmers have with programming the robot to retrieve the energy supply without detonating the bomb. *See id.* at 129–30.

²¹⁵ I state this is a likely flaw because, theoretically speaking, the violation of traffic laws may not necessarily be bad when it increases overall happiness.

²¹⁶ *See, e.g.*, D.C. CODE § 50-2352(3) (2015) (“An autonomous vehicle may operate on a public roadway; provided, that the vehicle . . . [i]s capable of operating in compliance with the District’s applicable traffic laws and motor vehicle laws and traffic control devices.”); FLA. STAT. ANN. § 319.145(1)(d) (West 2015) (“The vehicle shall . . . [b]e capable of being operated in compliance with the applicable traffic and motor vehicle laws of this state.”); NEV. ADMIN. CODE § 482A.190(2)(f) (2015) (“A certificate of compliance . . . must certify that the autonomous technology installed on the autonomous vehicle . . . [i]s capable of being operated in compliance with the applicable traffic laws of this State.”). Although the Michigan statute does not specifically require that the autonomous vehicle be capable of operating in compliance with its state’s traffic laws, that prohibition is presumed through its punishment scheme for violating laws. *See* MICH. COMP. LAWS ANN. §§ 257.663, 257.665, 257.666(2) (West 2015) (indicating that a person can be punished for violations of laws caused by the autonomous vehicle).

respectively.²¹⁷ Assume for example, that an autonomous vehicle encounters an (unrealistic) situation in which it must decide between crashing into and killing a human or crashing into and destroying a Boeing 787 Dreamliner (“Boeing”). The Boeing is priced at \$225 million, but sells on average for \$116 million.²¹⁸ Therefore, if the choice was destruction of the Boeing or the death of a person, the act-utilitarian autonomous vehicle would kill the person. Furthermore, the autonomous vehicle would kill the person even if the risk of destroying the Boeing was only ten percent. It seems wrong to value property over human life—or even to place a value on human life itself. A Boeing can be rebuilt—a human life cannot be reborn. A major flaw of an act-utilitarian autonomous vehicle is that it may not be able to take into account that simple fact.

B. Deontology (Kantianism)

For those people who consider a life priceless, they may find solace in a deontology theory of ethics; under which, the rightness of an act is based on the means or duties.²¹⁹ There are many types of deontology ethical theories, including W.D. Ross’s *prima facie* duties, Asimov’s laws of robotics, the Ten Commandments, and Kantianism.²²⁰ This Part focuses on the theory of ethics created by the most famous deontologist, Immanuel Kant.

For Kant, the motives of the underlying action matter.²²¹ As such, the consequences of an action do not determine its rightness: some actions are morally impermissible regardless of their consequences. Thus, some morally right actions have bad consequences, and some morally wrong actions have good consequences. For example, Kant reasoned that it is always morally wrong to tell a lie.²²² This principal can be illustrated through an example where a good friend comes to

²¹⁷ Binyamin Appelbaum, *As U.S. Agencies Put More Value on a Life, Businesses Fret*, N.Y. TIMES (Feb. 16, 2011), <http://www.nytimes.com/2011/02/17/business/economy/17regulation.html>.

²¹⁸ Agustino Fontevecchia, *Boeing Bleeding Cash as 787 Dreamliners Cost \$200M but Sell for \$116M, but Productivity is Improving*, FORBES (May 21, 2013), <http://www.forbes.com/sites/afontevecchia/2013/05/21/boeing-bleeding-cash-as-787-dreamliners-cost-200m-but-sell-for-116m-but-productivity-is-improving/>.

²¹⁹ See POJMAN, *supra* note 173, at 131–32; Thomas M. Powers, *Prospects for a Kantian Machine*, IEEE INTELLIGENT SYS., July–Aug. 2006, at 46, 46.

²²⁰ See, e.g., Gips, *supra* note 196, at 247, 248 (“There are many examples of deontological moral systems that have been proposed.”).

²²¹ See Robert Johnson, *Kant’s Moral Philosophy*, STAN. ENCYC. PHIL., <http://plato.stanford.edu/entries/kant-moral/> (last updated Apr. 6, 2008).

²²² See, e.g., RACHELS & RACHELS, *supra* note 177, at 127 (noting that Kant said it was always immoral to tell a lie).

your house after a fight with his wife.²²³ An hour later, his wife shows up at your door; she is angry, has a gun, and is looking for her husband. Kant would insist that you tell the truth about your friend being at your house, regardless of what that revelation might entail.²²⁴

Kant used absolute rules—better known as “categorical imperatives”—to formulate his theory of ethics.²²⁵ Kant deduced three such formulations of the categorical imperative. The first categorical imperative (“CI(1)”) states: “I ought never to act except in such a way *that I can also will that my maxim should become a universal law*.”²²⁶ CI(1) is a procedure by which a rational agent determines whether an act is morally permissible: an action is morally permissible only if an agent could consistently will that everyone engage in that action.²²⁷ Returning to the lying example, Kant stated that someone could not will everyone to lie because such a rule would be a self-defeating law: no one would believe each other once lying became commonplace and that would cause people to stop paying attention to what anyone said.²²⁸ Accordingly, lying could not be willed into a universal law, and, thus, lying is always immoral.²²⁹

Kant’s second formation of the categorical imperative (“CI(2)”) states: “Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end.”²³⁰ CI(2) only prohibits people from treating others *merely* as a means to their end; Kant had no “general objection to using people, or to using them as a means.”²³¹ In many non-objectionable situations, people use others as means.²³² Rather, CI(2) is based on Kant’s belief that each person

²²³ This example is a variation of an example provided in NILS CH. RAUHUT, *ULTIMATE QUESTIONS: THINKING ABOUT PHILOSOPHY* 223 (3d ed. 2011). See Tim Stelzig, Comment, *Deontology, Governmental Action, and the Distributive Exemption: How the Trolley Problem Shapes the Relationship Between Rights and Policy*, 146 U. PA. L. REV. 901, 907 (1998).

²²⁴ IMMANUEL KANT, *On a Supposed Right to Lie from Altruistic Motives*, in *CRITIQUE OF PRACTICAL REASON AND OTHER WRITINGS IN MORAL PHILOSOPHY* 346, 347 (Lewis White Beck ed., trans., 1949) [hereinafter KANT, *On a Supposed Right to Lie*].

²²⁵ RACHELS & RACHELS, *supra* note 177, at 128.

²²⁶ IMMANUEL KANT, *GROUNDWORK OF THE METAPHYSIC OF MORALS* 70 (H.J. Paton trans., 1964) [hereinafter KANT, *GROUNDWORK*].

²²⁷ See Powers, *supra* note 219, at 47 (“The first formulation of the categorical imperative supplies a procedure for producing ethical rules.”).

²²⁸ RACHELS & RACHELS, *supra* note 177, at 129.

²²⁹ *Id.*

²³⁰ KANT, *GROUNDWORK*, *supra* note 226, at 96.

²³¹ R. George Wright, *Treating Persons as Ends in Themselves: The Legal Implications of a Kantian Principle*, 36 U. RICH. L. REV. 271, 271, 277 (2002).

²³² See, e.g., FELDMAN, *supra* note 183, at 121 (“[W]e use other people as means to our ends all the time, and we cannot avoid doing so.”); Wright, *supra* note 231, at 277; Johnson, *supra*

has inherent value and must be treated with dignity.²³³ Thus, under CI(2), people should be treated differently than inanimate objects.²³⁴ Revisiting the Boeing example from the utilitarian subsection, it is clear that a Kantian autonomous vehicle would be programmed to value human life over property. As such, a Kantian autonomous vehicle would always destroy property rather than kill or even harm a person, regardless of the property's value.

Kant's final formulation of the categorical imperative ("CI(3)") states: "the Idea of the will of every rational being as a will which makes universal law."²³⁵ CI(3) represents Kant's belief that an act must be done out of a sense of duty—and not out of inclination—for the act to have moral worth.²³⁶ Acts done out of inclination are morally worthless.²³⁷

Autonomous vehicles will lack the capability to will their decisions as universal laws.²³⁸ The vehicle is also incapable of acting autonomously in accordance with CI(3). Therefore, an autonomous vehicle will not be a "moral agent" according to Kantian ethics.²³⁹ However, the programmer of the crash-optimization algorithm can

note 221 ("First, the Humanity formula does not rule out using people as means to our ends. Clearly this would be an absurd demand, since we do this all the time. Indeed, it is hard to imagine any life that is recognizably human without the use of others in pursuit of our goals.").

²³³ POJMAN, *supra* note 173, at 145.

²³⁴ FELDMAN, *supra* note 183, at 121.

²³⁵ KANT, GROUNDWORK, *supra* note 226, at 98.

²³⁶ FELDMAN, *supra* note 183, at 128.

²³⁷ *Id.*

²³⁸ See Gips, *supra* note 196, at 249 ("The categorical imperative might be tough to implement in a reasoning system."). For an artificial moral agent to make this decision, it: would need an explicit and fully stated principle of practical reason consisting of three elements: a goal, a means or course of action by which the agent proposes to achieve that goal, and a statement of the circumstances under which acting in that way will achieve the goal in question. Given these three elements, a very powerful computing device might be able to run an analysis or a simulation model to determine whether its goal would be blocked if all other agents were to operate with the same maxim.

WALLACH & ALLEN, *supra* note 12, at 95. In addition, the artificial moral agent would need to know about human psychology and the effects of its actions in the world. *Id.* at 96.

²³⁹ See, e.g., Ryan Tonkens, *A Challenge for Machine Ethics*, 19 MINDS & MACHINES 421, 422–23 (2009) (asserting that not only are artificial moral agents incapable of being Kantian moral agents, but the creation of those agents itself violates Kantian moral ethics); Anthony F. Beavers, Eighteenth Annual Meeting of the Ass'n for Practical & Prof'l Ethics: Between Angels and Animals: The Question of Robot Ethics, or is Kantian Moral Agency Desirable 3 (Mar. 2009) faculty.evansville.edu/tb2/PDFs/Robot%20Ethics%20-%20APPE.pdf ("The 'moral agent' in Kant's picture is the person suspended between inclination and duty, where inclination is derived from desire which, in turn, is always fixed to something outside the self and where duty is determined by pure practical reason as action in conformity with the 'universality of a law as such.' This suspension, to be clear, does not make the person moral, but rather *able to be* moral . . .") (citation omitted); *Id.* at 11 ("[T]o produce a robot that is a genuine moral agent, we must design within it the capacity to act immorally. Doing so, it seems to [the author], would be straight out immoral . . .").

act autonomously and will her programming decisions as universal laws, making her a rational agent. Thus, an autonomous vehicle could be programmed according to Kantian ethics.

In the Shopping Cart Problem from Part III, a Kantian autonomous vehicle would not strike the baby carriage.²⁴⁰ Unlike the shopping cart or the grocery store, the baby has inherent value that must be respected. Striking the baby carriage and killing the baby to cause less damage to the autonomous vehicle would be treating the baby as merely a means to the ends of the autonomous vehicle and its occupant. Therefore, the Kantian autonomous vehicle would not hit the baby carriage.

In the Trolley Problem, the Kantian autonomous vehicle must decide whether to kill one person, or let its human operator kill five persons. Initially, one could assert, perhaps plausibly, that killing is killing.²⁴¹ Under CI(1), killing could not be universalized as a universal law: if killing was universalized, everyone would be dead and no one could kill anyone else.²⁴² Therefore, killing violates CI(1). In addition to violating CI(1), the decision to take control of the vehicle to kill the person may violate CI(2). The issue is whether killing the one person to save the five persons is considered treating the one person as merely a means to the ends of the five persons.

Some philosophers have used the “doctrine of the double effect” to distinguish between foreseeable killings and intentional killings.²⁴³ “The doctrine of the double effect is based on a distinction between what a man foresees as a result of his voluntary action and what, in the strict sense, he intends.”²⁴⁴ Under this doctrine, the incidental killing of one person to save five persons is morally defensible but the affirmative killing of one person to save five persons is not.²⁴⁵ An example of the latter situation (Transplant Problem) is this:

David is a great transplant surgeon. Five of his patients need new parts—one needs a heart, the others need, respectively, liver, stomach, spleen, and spinal

²⁴⁰ See *supra* Section III.A; notes 233–34 and accompanying text.

²⁴¹ See *Why Ethics Matters*, *supra* note 14, at 79 (stating that a non-consequentialist might object to killing one person).

²⁴² See *supra* notes 226–27 and accompanying text.

²⁴³ See, e.g., Foot, *supra* note 144, at 5–6. The Doctrine of the Double Effect was introduced by Thomas Aquinas in his discussion of self-defense killing. See Alison McIntyre, *Doctrine of Double Effect*, STAN. ENCYC. PHIL., <http://plato.stanford.edu/entries/double-effect/> (last updated Sept. 23, 2014).

²⁴⁴ Foot, *supra* note 144, at 5–6.

²⁴⁵ See Judith Jarvis Thomson, *Killing, Letting Die, and the Trolley Problem*, MONIST, Apr. 1976 at 204, 206–07.

cord—but all are of the same, relatively rare, blood-type. By chance, David learns of a healthy specimen with that very blood-type. David can take the healthy specimen's parts, killing him, and install them in his patients, saving them. Or he can refrain from taking the healthy specimen's parts, letting his patients die.²⁴⁶

From a purely act-utilitarian standpoint, there is no difference between the Trolley Problem and the Transplant Problem. In both cases, five persons are saved at the expense of one person—five persons have more utility than one person. However, it is clear in the Transplant Problem that David is using the healthy specimen merely as a means to save the other five persons—he is cutting that person up so the others could live.²⁴⁷ Therefore, the Transplant Problem clearly violates CI(2). But the two problems seem different, even though the question is whether to sacrifice one person for the sake of five persons.²⁴⁸ In the Trolley Problem, the autonomous technology is not intending to kill the one person; instead, the car is intending to save the five persons, and the death of the one person is an incidental result of that intention. Nonetheless, as this discussion has made evident, resolution of the Trolley Problem in Kantian ethics is murkier than it was in utilitarianism.

This raises another concern with the application of Kantian ethics to autonomous vehicles: Kantianism cannot address every situation confronting an autonomous vehicle.²⁴⁹ In the Tunnel Problem, the autonomous vehicle must decide whether to kill a child or the operator of the vehicle. As indicated earlier, CI(1) prohibits killing; in either situation, someone dies. Thus, Kantianism fails to provide a rule of action. Likewise, in the Motorcycle Problem and the Car Problem, Kantian ethics does not provide an answer for which motorcycle or car to hit. In addition to that limitation, another problem for a deontological approach to autonomous vehicles is that

²⁴⁶ *Id.* at 206.

²⁴⁷ See Thomson, *supra* note 146, at 1401 (“It is striking, after all, that the surgeon who proceeds in *Transplant* treats the young man he cuts up ‘as a means only’: He literally uses the young man’s body to save his five, and does so without the young man’s consent.”).

²⁴⁸ See Foot, *supra* note 144, at 8 (“The doctrine of the double effect . . . insist[s] that it is one thing to steer towards someone foreseeing that you will kill him and another to aim at his death as part of your plan.”).

²⁴⁹ See, e.g., Goodall, *Ethical Decision Making*, *supra* note 1, at 62 (“Although deontological ethics can provide guidance in many situations, it is not suitable as a complete ethical system because of the incompleteness of any set of rules and the difficulty involved in the articulation of complex human ethics as a set of rules.”).

multi-rule systems suffer from conflicts between rules.²⁵⁰ A Kantian autonomous vehicle would also suffer from a frame problem: an algorithm writer would need to make sure the rules are activated when an autonomous vehicle encounters the situation necessitating its application.²⁵¹ Therefore, although Kantianism can provide some moral rules that could be applied to an autonomous vehicle, it could not be the only moral theory programmed into the vehicle.

V. LEGAL CONCERNS

Accidents caused by autonomous vehicles will likely implicate tort and criminal law.²⁵² Tort law is concerned with vindicating the victim and also “to confirm and reinforce public standards of behavior.”²⁵³ Criminal law focuses on vindicating the “state’s interests in deterring crime and imposing justice.”²⁵⁴ Or, perhaps more simply stated, “tort law prices, while criminal law prohibits.”²⁵⁵

Because tort and criminal law draw a distinction between action and inaction in the law, this Part starts by addressing that distinction. This Part then examines the application of tort and criminal law to decisions made by a crash-optimization algorithm. As to tort law, this Part analyzes these decisions in relation to intentional torts, negligence, products liability, and punitive damages. As to criminal law, this Part looks at criminal law more generally, focusing on concepts of intent and necessity.

A. *The Non-Duty to Act*

At common law, a person did not have “a duty to aid a stranger in distress even though the danger may be great and the inconvenience to the potential rescuer only slight.”²⁵⁶ In criminal law, most crimes

²⁵⁰ See, e.g., WALLACH & ALLEN, *supra* note 12, at 93; Gips, *supra* note 196, at 247 (“Whenever a multi-rule system is proposed, there is the possibility of conflict between the rules.”).

²⁵¹ See WALLACH & ALLEN, *supra* note 12, at 86 (“The designer of a deontological (ro)bot needs to find ways to ensure that the rules are activated when the situation requires their application . . .”).

²⁵² Frank Douma & Sarah Aue Palodichuk, *Criminal Liability Issues Created by Autonomous Vehicles*, 52 SANTA CLARA L. REV. 1157, 1161 (2012) (“Accidents that result in property damage or personal harm currently can have both criminal and civil implications for the driver at fault . . .”).

²⁵³ 1 DAN B. DOBBS ET AL., *THE LAW OF TORTS* §4, at 6 (2d ed. 2011).

²⁵⁴ *Id.*

²⁵⁵ John C. Coffee, Jr., *Does “Unlawful” Mean “Criminal”? Reflections on the Disappearing Tort/Crime Distinction in American Law*, 71 B.U. L. REV. 193, 194 (1991).

²⁵⁶ Claire Elaine Radcliffe, *A Duty to Rescue: The Good, the Bad and the Indifferent—The*

require affirmative action and not merely a failure to act,²⁵⁷ “even when that aid can be rendered without danger or inconvenience.”²⁵⁸ In tort law, “[t]he fact that the actor realizes or should realize that action on his part is necessary for another’s aid or protection does not of itself impose upon him a duty to take such action.”²⁵⁹ Therefore, “[a]bsent special relationships or particular circumstances or actions, a defendant is not liable in tort for a pure failure to act for the plaintiff’s benefit.”²⁶⁰

A cogent example of this non-duty to act is the tragic death of Catherine “Kitty” Genovese. A man followed Kitty home and stabbed her to death in the streets of her middle-class neighborhood for over thirty-five minutes.²⁶¹ Many neighbors peered through their curtains and saw the attack, but they did not call the police or intervene.²⁶² From a moral standpoint, the decision not to help Kitty was morally wrong—at the very least, her neighbors could have called the police.²⁶³ A moral duty to act, however, is not enough to impute criminal or tort liability.²⁶⁴ The bystanders “were legally unrepachable.”²⁶⁵

The Trolley Problem from Part III presents an interesting application of these rules.²⁶⁶ In the Trolley Problem, the vehicle was not in autonomous mode, meaning that the human driver would have

Bystander’s Dilemma, 13 PEPP. L. REV. 387, 389 (1986); see also Goodall, Presentation, *supra* note 50, at 3 (“In general, there is no duty to act affirmatively according to common law.”).

²⁵⁷ See, e.g., 1 WAYNE R. LAFAVE, SUBSTANTIVE CRIMINAL LAW § 6.2, at 434 (2d ed. 2003) (“Most crimes are committed by affirmative action rather than by non-action.”).

²⁵⁸ *Id.* § 6.2(a), at 436.

²⁵⁹ RESTATEMENT (SECOND) OF TORTS § 314 (AM. LAW INST. 1965).

²⁶⁰ 2 DAN B. DOBBS ET AL., THE LAW OF TORTS § 405, at 651 (2d ed. 2011). The relationships that impose duties include when a defendant has caused harm to the plaintiff; the plaintiff and the defendant have a special relationship; the defendant starts to offer assistance; and when the defendant assumes a duty. *Id.* § 405, at 654.

²⁶¹ Diane Kiesel, *Who Saw This Happen?: States Move to Make Crime Bystanders Responsible*, A.B.A. J., Sept. 1983, at 1208, 1208 (1983).

²⁶² Kiesel, *supra* note 261. The first person to call the police claimed he waited so long because he did not “want to get involved.” *Id.*

²⁶³ See, e.g., POJMAN, *supra* note 173, at 114 (“We should call the police or do what is necessary to help her (as long as helping her does not create more disutility than leaving her alone).”); Ernest J. Weinrib, *The Case for a Duty to Rescue*, 90 YALE L.J. 247, 279–92 (1980) (discussing the philosophical arguments for rescuing).

²⁶⁴ See, e.g., 1 LAFAVE, *supra* note 257, § 6.2(a), at 436 (“A moral duty to take affirmative action is not enough to impose a legal duty to do so.”); Weinrib, *supra* note 263, at 258 (“Both courts and commentators generally consider it morally outrageous that the defense of nonfeasance can deny endangered persons a legal right to easy rescue.”).

²⁶⁵ Radcliffe, *supra* note 256, at 387; see also Tom Stacy, *Acts, Omissions, and the Necessity of Killing Innocents*, 29 AM. J. CRIM. L. 481, 514 (2002) (“Under the general American rule that an actor has no legal responsibility for the consequence of inaction, none of the onlookers has any criminal liability.”).

²⁶⁶ See *supra* Section III.F.

killed five people. Although many scholars have argued that autonomous vehicles should be civilly liable for accidents caused in autonomous mode,²⁶⁷ no one has argued that car manufacturers should be liable for accidents caused when a person is driving the car. Therefore, if the autonomous technology lets the five people die, then the car manufacturer is not liable for the accident.²⁶⁸ However, when the autonomous technology takes control over the vehicle to save the five people, it causes the death of the one person.²⁶⁹ Therefore, by doing what many would consider right,²⁷⁰ the car manufacturer would subject itself to legal liability.

A more troublesome application of this doctrine would occur if the autonomous technology could take control over the vehicle to save five persons but the accident does not result in a death; assume instead, the vehicle would crash into a parked vehicle. In this variation, the car manufacturer would likely not be liable for the death of the five people, but it would be liable for the damage to the parked vehicle if the autonomous technology took control over the vehicle.²⁷¹ Therefore, a car manufacturer may not be incentivized to use those accident-avoidance mechanisms and crash-optimization algorithms that would subject it to liability.

B. Tort Law

When autonomous vehicles cause accidents, the aggrieved parties will likely sue the car manufacturers, operators, or both for their injuries and damage to their property.²⁷² As control of the major operations of the vehicle transfers from the driver to the autonomous technology, it seems less likely that the traditional driver will be responsible for an accident caused by an autonomous vehicle in

²⁶⁷ See, e.g., *Why Ethics Matters*, *supra* note 14, at 79.

²⁶⁸ *Id.* (“If the car does not wrestle control from the human driver, then it (and the OEM) would perhaps not be responsible for the deaths of the five pedestrians while you were driving the car; it is merely letting those victims die.”).

²⁶⁹ *Id.* (“[I]f the car does take control and make a decision that results in the death of a person, then it (and the OEM) becomes responsible for killing a person.”).

²⁷⁰ See, e.g., Sarah Bakewell, *Clang Went the Trolley: ‘Would you Kill the Fat Man?’ and ‘The Trolley Problem,’* N.Y. TIMES (Nov. 22, 2013), <http://www.nytimes.com/2013/11/24/books/review/would-you-kill-the-fat-man-and-the-trolley-problem.html> (“Most people say . . . [b]etter that one person should die instead of five”).

²⁷¹ “Likely” is used because as indicated in the design defect part of products liability, the product may be defective for failing to save five lives when the autonomous technology could have saved them in this situation. See *infra* Section V.B.3.

²⁷² See Colonna, *supra* note 26, at 102 (“[W]henver autonomous technology is controlling a means of transportation and causes harm or damage, the plaintiffs bring products or strict liability claims against the manufacturer.”).

autonomous mode.²⁷³ As I have argued elsewhere, and many others have, it is likely that tort liability will be imposed on the manufacturer of an autonomous vehicle for accidents caused by its autonomous technology.²⁷⁴ This subsection will operate under the assumption that the manufacturer is liable.

This subsection analyzes the application of tort law to accidents involving the use of crash-optimization algorithms.²⁷⁵ When a tortfeasor harms someone intentionally, the tortfeasor is typically sued for an intentional tort.²⁷⁶ When the tortfeasor acts unintentionally but unreasonably, the remedy for the aggrieved party is in negligence.²⁷⁷ Special issues arise when a person is injured by a product; when that occurs, the aggrieved party typically sues the manufacturer of the product under theories of products liability.²⁷⁸ This subsection concludes by examining punitive damages—an overriding concern for all tortfeasors.

1. Intentional Torts

Because the decisions made by a crash-optimization algorithm are pre-programmed into the vehicle, those harmed by a crash-

²⁷³ See, e.g., F. Patrick Hubbard, “Sophisticated Robots”: *Balancing Liability, Regulation, and Innovation*, 66 FLA. L. REV. 1803, 1858–59 (2014) (“It is likely that, as automobiles become increasingly robotic, accidents will be caused more and more by features of the car itself and less and less by the conduct of drivers. This increased role of the car vis-à-vis the driver will likely shift liability costs away from owners and drivers, who could be liable based on negligence law, to sellers and distributors, who could be liable under products liability law for defective features.”); ANDERSON ET AL., *supra* note 60, at 115.

²⁷⁴ See *infra* note 275.

²⁷⁵ For a discussion of tort liability’s application to autonomous vehicles more generally, see Jack Boeglin, *The Costs of Self-Driving Cars: Reconciling Freedom and Privacy with Tort Liability in Autonomous Vehicle Regulation*, 17 YALE J. L. & TECH. 171, 175 (2015); Duffy & Hopkins, *supra* note 26, at 456; Dorothy J. Glancy, *Sharing the Road: Smart Transportation Infrastructure*, 41 FORDHAM URB. L.J. 1617, 1643–47 (2014); Kyle Graham, *Of Frightened Horses and Autonomous Vehicles: Tort Law and its Assimilation of Innovations*, 52 SANTA CLARA L. REV. 1241, 1241–42 (2012); Hubbard, *supra* note 273, at 1806; Gary E. Marchant & Rachel A. Lindor, *The Coming Collision Between Autonomous Vehicles and the Liability System*, 52 SANTA CLARA L. REV. 1321, 1322 (2012); Vladeck, *supra* note 6, at 127–28; Colonna, *supra* note 26, at 104; Kevin Funkhouser, Note, *Paving the Road Ahead: Autonomous Vehicles, Products Liability, and the Need for a New Approach*, 2013 UTAH L. REV. 437, 440 (2013); Garza, *supra* note 18, at 583; Jeffrey K. Gurney, Note, *Sue My Car Not Me: Products Liability and Accidents Involving Autonomous Vehicles*, 2013 U. ILL. J.L. TECH. & POL’Y 247, 251–52 (2013); Dylan LeValley, Note, *Autonomous Vehicle Liability—Application of Common Carrier Liability*, 36 SEATTLE U. L. REV. SUPRA 5, 6 (2013), <http://seattleuniversitylawreview.com/files/2013/02/SUPraLeValley2.pdf>.

²⁷⁶ See, e.g., Ugo Pagallo, *Three Roads to Complexity, AI and the Law of Robots: On Crimes, Contracts, and Torts*, in *AI APPROACHES TO THE COMPLEXITY OF LEGAL SYSTEMS* 48, 55 (Monica Palmirani et al. eds., 2012).

²⁷⁷ See, e.g., *id.*

²⁷⁸ See, e.g., *id.*

optimization algorithm may believe they were “targeted.”²⁷⁹ For example, in the Motorcycle Problem, if the vehicle is programmed to hit the motorcyclist with a helmet, a helmeted motorist may believe that the vehicle “intentionally” hit her because of her helmet use. The person who believes she was targeted may seek recourse through an intentional tort action.²⁸⁰

The distinguishing feature of an intentional tort is intent on the part of the tortfeasor. A tortfeasor has the intent to commit an intentional tort if she (1) desires the consequences of the act or (2) is substantially certain that the consequences will result from the act.²⁸¹ The first method of satisfying intent is the common way people think about intent, and it clearly embodies wrongfulness.²⁸² An example would be when a tortfeasor aims her gun at and shoots a person. The second method of establishing intent is more convoluted.²⁸³ A commonly used example of substantial certainty involves collateral harm²⁸⁴: A sets up a bomb in B’s office for the purpose of killing B, but A also knows that C is in the office and will die too, even though A does not intend to hurt C; nonetheless, A still detonates the bomb.²⁸⁵ In that example, C’s estate could sue A for the intentional tort of battery because A was substantially certain that C would die.²⁸⁶

A “crafty” plaintiff could assert that the algorithm writer commits the intentional tort of battery in the moral dilemmas from Part III.²⁸⁷ For example, in the Trolley Problem, the estate of the person killed could assert that the car manufacturer programmed its vehicle with

²⁷⁹ See *Why Ethics Matters*, *supra* note 14, at 73.

²⁸⁰ See, e.g., RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL & EMOTIONAL HARM § 1 cmt. a (AM. LAW INST. 2010) (explaining that liability in tort will often turn on whether and to what extent the requisite intent is present); *Why Ethics Matters*, *supra* note 14, at 73 (noting that when faced with an imminent crash, a good autonomous car would be programmed to target a motorcyclist with a helmet as opposed to one without a helmet).

²⁸¹ See RESTATEMENT (SECOND) TORTS § 8A (AM. LAW INST. 1965); 1 DOBBS ET AL., *supra* note 253, § 29, at 73.

²⁸² See, e.g., RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL & EMOTIONAL HARM § 1 cmt. a (AM. LAW INST. 2010) (“There is a clear element of wrongfulness in conduct whose very purpose is to cause harm.”).

²⁸³ See, e.g., *id.* (“[T]here are complications in considering the liability implications of harms that are intentional only in the sense that the actor who engages in conduct knows that harm is substantially certain to result.”).

²⁸⁴ See, e.g., RESTATEMENT (SECOND) OF TORTS § 8A cmt. b, illus. 1.

²⁸⁵ *Id.*

²⁸⁶ See RESTATEMENT (SECOND) OF TORTS § 13 (AM. LAW INST. 1965) (“An actor is subject to liability to another for battery if he acts *intending* to cause a harmful . . . contact with . . . a third person . . . and a harmful contact with the person of the other . . . results.”) (emphasis added).

²⁸⁷ See discussion *supra* Part III.

the intent to kill that person, or more likely, that the car manufacturer was substantially certain that its car would kill someone when doing so would result in more lives saved. Under the first method of establishing intent, it seems unlikely that a car manufacturer intended the harm; the programming of the vehicle and the actual accident is too attenuated in time and space to impute intent.

Under the substantial certainty test, imposing liability also seems too attenuated.²⁸⁸ The comments to the Restatement (Third) of Torts state that the substantial certainty test is limited to:

[H]arm to a particular victim, or to someone within a small class of potential victims within a localized area. The test loses its persuasiveness when the identity of potential victims becomes vaguer and when, in a related way, the time frame involving the actor's conduct expands and the causal sequence connecting conduct and harm becomes more complex.²⁸⁹

An accident caused by a crash-optimization algorithm seems “too broad and unfocused to support liability based on intent”²⁹⁰—the algorithm writer did not know who would be harmed or when such harm would occur. Moreover, the manufacturer programmed the crash-optimization algorithm to mitigate harm and damage, and certainly if the manufacturer had its choice, no one would ever be harmed by its car.²⁹¹ Therefore, because the identity of the individual victim of a decision made by a crash-optimization algorithm will not be known when the algorithm is programmed, and because the certainty of harm is “at some undefined time and place,”²⁹² courts should not find the car manufacturer liable for an intentional tort. Instead, other tort doctrines better encompass injuries caused by the decisions of a crash-optimization algorithm.

²⁸⁸ A substantial certainty argument in this situation is no different than finding a baseball player responsible for an intentional tort when he hits a foul ball into the stands. See James A. Henderson, Jr. & Aaron D. Twerski, *Intent and Recklessness in Tort: The Practical Craft of Restating Law*, 54 VAND. L. REV. 1133, 1141–42 (2001). Certainly, every ball player's decision to swing the bat is an “intentional act,” and during the course of the season, a player knows that he will hit at least one foul ball into the stands. See *id.* However, no one seriously contends that a ballplayer commits an intentional tort every time a foul ball hits a fan. See *id.*

²⁸⁹ RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL & EMOTIONAL HARM § 1 cmt. e (AM. LAW INST. 2010).

²⁹⁰ Henderson & Twerski, *supra* note 288, at 1142 n.44.

²⁹¹ See *Why Ethics Matters*, *supra* note 14, at 72.

²⁹² See RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL & EMOTIONAL HARM § 1 cmt. e (AM. LAW INST. 2010).

2. Negligence

Liability for most car accidents is analyzed based on principles of negligence,²⁹³ which makes people liable for “*unreasonably* failing to prevent [a] risk.”²⁹⁴ To prove a negligence case, a plaintiff must show: (1) duty, (2) breach, (3) causation, and (4) damages.²⁹⁵ A defendant must have a duty before she can be liable for injuries.²⁹⁶ Generally, a defendant has a duty to exercise “reasonable care.”²⁹⁷ Another test used for determining the amount of care a defendant owed to a plaintiff is the “calculus of risk” test.²⁹⁸ Judge Learned Hand developed a famous formula for the calculus of risk test:

[A defendant’s duty of care] to provide against resulting injuries is a function of three variables: (1) The probability [of injury]; (2) the gravity of the resulting injury . . . ; (3) the burden of adequate precautions. Possibly it serves to bring this notion into relief to state it in algebraic terms: if the probability be called P; the injury, L; and the burden, B; liability depends upon whether B is less than L multiplied by P: i.e., whether B less than PL.²⁹⁹

A defendant breaches her duty when she fails to exercise reasonable care or when she engages in unreasonably risky conduct under the Hand formula.³⁰⁰ Causation has two parts: cause-in-fact and proximate cause.³⁰¹ Generally, a defendant is a cause-in-fact of the harm to the plaintiff when that harm would not have occurred “but for” defendant breaching her duty.³⁰² Historically, courts have

²⁹³ See 1 DOBBS ET AL., *supra* note 253, § 120, at 373–74 (indicating that automobile cases are typically based on negligence). The mandatory car insurance requirement has impacted the application of negligence to automobile accidents because insurers have informally adopted rules to allocate fault in an accident. See ANDERSON ET AL., *supra* note 60, at 113.

²⁹⁴ ANDERSON ET AL., *supra* note 60, at 112.

²⁹⁵ See 1 DOBBS ET AL., *supra* note 253, § 124, at 389.

²⁹⁶ See 2 DOBBS ET AL., *supra* note 260, § 251, at 2.

²⁹⁷ *Id.*; see also RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL & EMOTIONAL HARM § 7(a) (“An actor ordinarily has a duty to exercise reasonable care when the actor’s conduct creates a risk of physical harm.”).

²⁹⁸ F. PATRICK HUBBARD & ROBERT L. FELIX, THE SOUTH CAROLINA LAW OF TORTS 63 (4th ed. 2011).

²⁹⁹ *United States v. Carroll Towing Co.*, 159 F.2d 169, 173 (2d Cir. 1947); see also HUBBARD & FELIX, *supra* note 298, at 63–64, 65 (expounding Judge Hand’s conception of the formula).

³⁰⁰ 1 DOBBS ET AL., *supra* note 253, § 125, at 390–91.

³⁰¹ *Id.* § 125, at 392.

³⁰² *Bramlette v. Charter-Med.-Columbia*, 393 S.E.2d 914, 916 (S.C. 1990) (“Causation in fact is proved by establishing the injury would not have occurred ‘but for’ the defendant’s negligence.” (citing *Hanselmann v. McCordle*, 267 S.E.2d 531, 533 (S.C. 1980))); see also *Hughes v. Children’s Clinic, P.A.*, 237 S.E.2d 753, 757 (S.C. 1977) (finding that plaintiff’s injuries would not have resulted but for the presence of a mirror on the premises).

used two tests to determine whether the defendant's breach of her duty was the "proximate cause" of the plaintiff's harm: (1) directness and (2) foreseeability.³⁰³ "The directness rule extends to all outcomes, even if not foreseeable, so long as they are the 'direct' result of the tortious conduct."³⁰⁴ Under the foreseeability test, the defendant is only responsible for foreseeable results of her action—not freakish and unforeseen outcomes.³⁰⁵ And finally, a plaintiff must prove that damages resulted from the defendant's negligence.³⁰⁶

These elements can be applied to an accident caused by a driver in a traditional vehicle. Assume that the driver was checking her cell phone and caused an accident when her vehicle crossed into oncoming traffic. In such a case, the driver has a duty to exercise reasonable care while driving. She breached that duty when she failed to pay attention to the road by checking her phone. The accident would not have occurred "but for" her using her phone; such an accident was foreseeable when she used her phone. Damages resulted from the accident. Thus, the driver is liable for the accident.

Negligence can also be applied to an accident that occurs because of a malfunction in the autonomous technology, and liability can be imposed on the car manufacturer for that harm. A car manufacturer has a "duty to exercise reasonable care to refrain from selling products that contain unreasonable risks of harm" to consumers and others who could be foreseeably harmed by its products.³⁰⁷ It is also likely that an autonomous car will have a greater duty of care than a traditional driver because of its ability to discover danger better than a human driver.³⁰⁸ A plaintiff cannot establish a breach by merely pointing out that a defect occurred—manufacturers only have to exercise reasonable but not perfect care.³⁰⁹ Instead, the plaintiff must prove that the product was defective, and it was the

³⁰³ See, e.g., Jane Stapleton, *Legal Cause: Cause-in-Fact and the Scope of Liability for Consequences* 54 VAND. L. REV. 941, 996 (2001). For example, in the famous case of *Palsgraf v. Long Island Railroad Co.*, the majority opinion dismissed the complaint because the harm to the plaintiff was not a foreseeable result of the acts of the employees of the defendant, while the dissent focused on the causal chain between the employee's conduct and the resultant harm. *Palsgraf v. Long Island R. Co.*, 162 N.E. 99, 101 (N.Y. 1928); *Id.* at 105 (Andrews, J., dissenting).

³⁰⁴ Stapleton, *supra* note 303, at 996.

³⁰⁵ *Id.*

³⁰⁶ See 1 DOBBS ET AL., *supra* note 253, § 124, at 389.

³⁰⁷ See 1 DAVID G. OWEN & MARY J. DAVIS, OWEN & DAVIS ON PRODUCT LIABILITY § 2:2 (4th ed. 2014); Glancy, *supra* note 275, at 1644 n.135 ("Negligence liability would be based on a product manufacturer's failure to exercise reasonable care in designing or building a product that causes reasonably foreseeable harm.").

³⁰⁸ See Vladeck, *supra* note 6, at 130.

³⁰⁹ 1 OWEN & DAVIS, *supra* note 307, § 2:3.

manufacturer's negligence that caused the defectiveness.³¹⁰ The plaintiff would then need to prove that the car manufacturer's breach was the "but for" and proximate cause of her harm.³¹¹ And then the plaintiff would need to establish damages.³¹²

The principals of negligence can be applied to the moral dilemmas from Part III. In the Shopping Cart Problem, an operator has a duty to inspect her brakes, and a manufacturer has a duty to provide brakes that work. One breached that duty when the brakes failed. Had the operator inspected her brakes or the manufacturer supplied brakes in working fashion, no harm would have resulted—whether to the shopping cart, baby, or grocery store. That harm was a direct result of the breach, and it is foreseeable that failing to inspect brakes or supplying faulty brakes can result in a car accident. And damages did in fact occur. Therefore, the plaintiff could recover for negligence.

In the Bridge Problem, the driver of the school bus was negligent. That driver has a duty to drive on her side of the road. That driver breached her duty by driving in the wrong lane. If an accident ensues between the school bus and the vehicle, that breach will be the cause-in-fact and proximate cause of the accident, and damages would result. The question is whether the autonomous vehicle has a duty to avoid the accident. Generally speaking, a driver has a duty to exercise reasonable care to avoid an impending collision.³¹³ Failure to exercise reasonable care can be contributory or comparative negligence, which are affirmative defenses that reduce or bar a plaintiff's recovery.³¹⁴ However, the autonomous vehicle would likely not be under a duty to commit self-sacrifice;³¹⁵ the duty to avoid an accident does not require the person to undertake an action that

³¹⁰ *Id.*; see also *Merrill v. Navegar, Inc.*, 28 P.3d 116, 124 (Cal. 2001) ("[U]nder . . . negligence . . . , to recover from a manufacturer, a plaintiff must prove that a defect caused injury . . . [and] that the defect in the product was due to negligence of the defendant.").

³¹¹ See *Bramlette v. Charter-Med.-Columbia*, 393 S.E.2d 914, 916 (S.C. 1990).

³¹² See 1 DOBBS ET AL., *supra* note 253, § 124, at 389.

³¹³ *W.E. Shipley, Negligence of Motorist Colliding with Vehicle Approaching in Wrong Lane*, 47 A.L.R. 2d 6, § 2 (1956) ("[A] driver who, while proceeding on his proper side of the road, meets and collides with another vehicle approaching from the opposite direction on the wrong side . . . is ordinarily held to the standard of care of the ordinary reasonable man 'in the circumstances.'"); see also *Rutter v. Gemmer*, 505 N.E.2d 1308, 1313 (Ill. App. Ct. 1987) ("One must still exercise due care to prevent injury to oneself and others and to avoid a collision with a driver who may be proceeding on the wrong side of the road." (citing *Balfour v. Citgo Petroleum Corp.*, 452 N.E.2d 46, 49 (Ill. App. Ct. 1983))).

³¹⁴ See *Shipley*, *supra* note 313, at § 3; Thomas R. Trenkner, *Modern Development of Comparative Negligence Doctrine Having Applicability to Negligence Actions Generally*, 78 A.L.R. 3d 388, § 2[b] (1977).

³¹⁵ See generally *Shipley*, *supra* note 313, at § 2 (inferring that if drivers are only obligated to exercise reasonable care, there is no way drivers would ever be under a duty to self-sacrifice because that is far from what we would expect a reasonable person to do).

would result in her death.³¹⁶ And thus, negligence would probably not be imposed on the autonomous vehicle for its decision to crash into the school bus.

However, in other situations where the autonomous vehicle is not faced with the choice of killing its operator, the vehicle would likely be under a duty to avoid the accident. If the autonomous vehicle acts reasonably in avoiding an accident caused by someone else, it is not negligent because it did not breach its duty. If, however, the autonomous vehicle makes a poor decision, two specialized tort doctrines may shield it from liability: (1) the emergency doctrine; and (2) the unavoidable accident situation.³¹⁷

The emergency doctrine “recognizes that a driver who, although driving in a prudent manner, is confronted with a sudden or unexpected event which leaves little or no time to apprehend a situation and act accordingly should not be subject to liability simply because another perhaps more prudent course of action was available.”³¹⁸ Thus, the doctrine applies when a driver made an unwise decision while avoiding the accident.³¹⁹ The underlying rationale “is that a person faced with an emergency which his conduct did not create or help to create is not guilty of negligence in the

³¹⁶ *Id.*

³¹⁷ See George L. Blum, *Instructions on “Unavoidable Accident,” “Mere Accident,” or the like, in Motor Vehicle Cases—Modern Cases*, 21 A.L.R. 5th 82, § 2[a] (1994); Jeffrey F. Ghent, *Modern Status of the Emergency Doctrine*, 10 A.L.R. 5th 680, § 2 (1993).

³¹⁸ *Lockhart v. List*, 665 A.2d 1176, 1180 (Pa. 1995). A sudden emergency has been defined by one court as: “(1) an unforeseen combination of circumstances which calls for immediate action; (2) a perplexing contingency or complication of circumstances; (3) a sudden or unexpected occasion for action, exigency, pressing necessity.” *Vasconez v. Mills*, 651 N.W.2d 48, 54 (Iowa 2002) (quoting *Foster v. Strutz*, 636 N.W.2d 104, 106 (Iowa 2001)). Courts have instructed juries on the emergency doctrine in various situations. See, e.g., *Stephens v. Hypes*, 610 S.E.2d 631, 633 (Ga. Ct. App. 2005) (“[C]ommon sense dictates that a deer in the road could give rise to a sudden emergency.”); *Cordas v. Peerless Transp. Co.*, 27 N.Y.S.2d 198, 199, 202 (1941) (finding sudden emergency when there was a physical threat to the driver); *Marsch v. Catanzaro*, 837 N.Y.S.2d 195, 197 (N.Y. App. Div. 2007) (“A driver is not obligated to anticipate that a vehicle traveling in the opposite direction will cross over into the oncoming lane of traffic. Such an event constitutes a classic emergency situation, implicating the emergency doctrine.” (quoting *Gajjar v. Shah*, 817 N.Y.S.2d 653, 654 (N.Y. App. Div. 2006))); *Rambo v. McCulloch*, 752 P.2d 347, 349 (Or. Ct. App. 1988) (finding emergency situation due to jammed brakes).

³¹⁹ Gregory M. Wasson, *Existence of “Sudden Emergency,”* 8 AM. JUR. PROOF OF FACTS 3D. 399, § 15 (1990) (“In order for the doctrine to be applicable, the circumstances surrounding the emergency must be such that the actor had an opportunity to choose the course of conduct which subsequently appears to have been unwise.”); see also *Dupree v. Sayes*, 974 So. 2d 22, 24 (La. Ct. App. 2007) (“Under the sudden emergency doctrine, one who finds himself in a position of imminent peril, without sufficient time to consider and weigh all the circumstances or the best means to adopt in order to avoid an impending danger, is not guilty of negligence if he fails to adopt what subsequently and upon reflection may appear to be the better method, unless the emergency is brought about by his own negligence.” (quoting *Smeby v. Williams*, 961 So. 2d 597, 600 (La. Ct. App. 2007))).

methods he chose, or failed to choose, to avoid the threatened disaster if he is compelled to act instantly without time for reflection.”³²⁰ Accordingly, “[a] person faced by such an emergency has some leeway when deciding rapidly between alternative courses of action.”³²¹ When properly invoked, “[t]he emergency doctrine excuses an individual from negligence.”³²²

Two limitations will hinder the application of this doctrine. First, the emergency doctrine does not shield a defendant’s liability when her negligence created the emergency.³²³ When a person confronts an emergency created by her own negligence, that person is liable based on the underlying negligence that created the emergency.³²⁴ For example, in the Shopping Cart Problem, the autonomous vehicle manufacturer or the operator could not use the emergency doctrine because one of them caused the “emergency.” Second, some states have abolished or restricted the use of the emergency doctrine.³²⁵

The other tort doctrine that could insulate a defendant from liability is the “unavoidable-accident doctrine.” This doctrine is used when “an occurrence or happening as, under all attendant circumstances and conditions, could not have been foreseen or anticipated in the exercise of ordinary care as the proximate cause of injury by any of the parties concerned.”³²⁶ Thus, it is limited to cases where neither party was negligent.³²⁷ Because neither party is at

³²⁰ *Totsky v. Riteway Bus Serv., Inc.*, 2000 WI App 29, ¶ 23, 233 Wis. 2d 371, 387, 607 N.W.2d 637, 644 (Wis. 2000) (quoting *Seif v. Turowski*, 181 N.W.2d 388, 392 (Wis. 1970)) (internal quotation marks omitted).

³²¹ *Oberempt v. Egri*, 410 A.2d 482, 484–85 (Conn. 1979) (citing *Mei v. Alterman Transp. Lines, Inc.*, 268 A.2d 639, 642 (Conn. 1970)).

³²² *Totsky*, 2000 WI App 29 at ¶ 22, 233 Wis. 2d at 387, 607 N.W.2d at 643 (citing *Seif*, 181 N.W.2d at 392).

³²³ *See, e.g., Mitchell v. Johnson*, 641 So. 2d 238, 239 (Ala. 1994) (“[O]ne who has by his or her own conduct brought about a sudden peril may not invoke as a defense the sudden emergency doctrine.” (citing *Glanton v. Huff*, 404 So. 2d 11, 13 (Ala. 1981))); *Seitner v. Clevenger*, 68 So. 2d 396, 396, 397 (Fla. 1953) (“One cannot defend on a theory of sudden emergency when his own negligent action brings it into existence. . . . To recognize the right of a defendant to escape liability under such circumstances would be to reward one for his own negligence.”); RESTATEMENT (SECOND) OF TORTS § 296 cmt. d (AM. LAW INST. 1965) (“Where the emergency itself has been created by the actor’s own negligence or other tortious conduct, the fact that he has then behaved in a manner entirely reasonable in the light of the situation with which he is confronted does not insulate his liability for his prior conduct.”).

³²⁴ RESTATEMENT (SECOND) OF TORTS § 296 cmt. d (AM. LAW INST. 1965).

³²⁵ *See, e.g., Wiles v. Webb*, 946 S.W.2d 685, 689 (Ark. 1997).

³²⁶ *Blum, supra* note 317, § 2[a].

³²⁷ *See Bashi v. Wodarz*, 53 Cal. Rptr. 2d 635, 639 (Cal. Ct. App. 1996) (“In effect, the concept of unavoidable accident is predicated upon absence of fault.”); *Blum, supra* note 317, § 11[a] (noting the applicability of the unavoidable accident doctrine when neither party was negligent in a motor vehicle accident (citing *Gwinn v. Payne*, 477 P.2d 680, 684 (Okla. 1970))); 57A AM. JUR. 2D *Negligence* § 38 (1994).

fault for the accident, there is no need for this doctrine: a defendant is not liable when she does not breach a duty to the plaintiff or is not the proximate cause of the accident.³²⁸

In the Bridge Problem, assume that instead of the school bus drifting into the wrong lane, a car full of college students tried to pass the school bus and did not see the oncoming autonomous vehicle. In that situation, the driver of the car breached her duty by not looking to see if it was safe to pass. The autonomous vehicle would have a duty to exercise reasonable care—i.e., not turn the vehicle into the school bus full of children. If the vehicle does turn into the school bus full of children, the defendant would want to use the emergency doctrine as a defense because it protects a defendant from an unwise decision. If, however, the autonomous vehicle stays the course and hits the car, it will be difficult to see how the autonomous vehicle was negligent for not driving off the bridge.

3. Products Liability

Products liability is a specialized form of tort liability that can be used by plaintiffs to sue a product manufacturer for injuries caused by that manufacturer's products.³²⁹ When an autonomous vehicle causes an accident because of a malfunction in the vehicle's technology, a plaintiff may assert that the accident was a result of a product defect.³³⁰ Generally, a plaintiff can bring three types of product liability claims: (1) manufacturing defects; (2) warning defects; and (3) design defects.³³¹ A manufacturing defect occurs when the product does not meet the manufacturer's design or specification.³³² A defect of this sort can be proven by comparing the product that harmed the plaintiff with the manufacturer's

³²⁸ See, e.g., *City of Phoenix v. Camfield*, 400 P.2d 115, 120 (Ariz. 1965) (en banc) ("If the jury is properly instructed on the necessity of negligence on the part of the defendant and the request that this negligence be the proximate cause of the injury that is sufficient. The defense of unavoidable accident is actually a defense of non negligence and an instruction on unavoidable accident is, in fact, confusing. It improperly implies that 'unavoidable accident' is a separate and distinct defense from 'non negligence.'"); *Dyer v. Herb Prout & Co.*, 498 A.2d 715, 717 (N.H. 1985) ("The instruction often sounds like a defense, but it reflects nothing more than a denial by the defendant of negligence or a claim that his negligence, if any, was not the proximate cause of the plaintiff's injuries." (citing *Butigan v. Yellow Cab Co.*, 320 P.2d 500, 504 (Cal. 1958))).

³²⁹ See Pagallo, *supra* note 276, at 55.

³³⁰ See, e.g., Gurney, *supra* note 275, at 258–59.

³³¹ RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2 (AM. LAW INST. 1998).

³³² See, e.g., David G. Owen, *Manufacturing Defects*, 53 S.C. L. REV. 851, 865 (2002) ("A defect in manufacture simply meant that through some mistake in the production process the product was rendered 'defective.'").

specifications for that product; the product has a manufacturing defect if it does not meet the specifications.³³³ Because this article is analyzing the purposeful decisions made by a crash-optimization algorithm, manufacturing defects are beyond the scope of this article.³³⁴

Likewise, issues revolving around liability for decisions made by a crash-optimization algorithm will unlikely be resolved as a warning defect. The warning defect doctrine requires manufacturers to provide consumers with two types of information: (1) notice of hidden dangers, and (2) instructions.³³⁵ Because of the nature of the decisions made by a crash-optimization algorithm, a manufacturer should provide some notice to the owner of the vehicle that, for example, the autonomous vehicle will not necessarily always protect the occupant of the vehicle—if that is true.³³⁶ Other than that, it is difficult to see how warning defects will have much application in these situations.

The majority of the product liability issues for an autonomous vehicle manufacturer will be design defects.³³⁷ Courts use two tests to determine whether a product has a design defect: (1) consumer-expectations test, and (2) risk-utility test.³³⁸ The consumer-expectations test was the initial defect test applied by courts; however, over time, many jurisdictions have abandoned this theory of design defectiveness.³³⁹ Because the consumer-expectations test has not been completely rejected,³⁴⁰ this subsection will analyze its

³³³ See *id.* at 866–70 (discussing the “departure-from-design test”).

³³⁴ For a discussion on autonomous vehicles and manufacturing defects, see Gurney *supra* note 275, at 258–60. However, even when an error or bug causes the underlying accident, it is unlikely that a plaintiff could prevail on a manufacturing defect because the software would cause the error. See *id.* at 259.

³³⁵ 1 OWEN & DAVIS, *supra* note 307, at § 9:1; see also RESTATEMENT (THIRD) TORTS: PRODS. LIAB. § 2(c) (AM. LAW INST. 1998) (“A product . . . is defective because of inadequate instructions or warnings when the foreseeable risks of harm posed by the product could have been reduced or avoided by the provision of reasonable instructions or warnings by the seller or other distributor, or a predecessor in the commercial chain of distribution, and the omission of the instructions or warnings renders the product not reasonably safe.”).

³³⁶ See Gurney, *supra* note 275, at 264.

³³⁷ See *id.* at 260 (maintaining that since malfunction doctrine limitations will render a traditional manufacturing defect claim useless, plaintiffs are more likely to assert a design defect claim).

³³⁸ David G. Owen, *Design Defects*, 73 MO. L. REV. 291, 299 (2008) (“All courts judge the adequacy of a product’s design upon one of two basic standards, or some combination thereof: (1) the ‘consumer expectations’ test—whether the design meets the safety expectations of users and consumers, and/or (2) the ‘risk-utility’ test—whether the safety benefits of designing away a foreseeable danger exceed the resulting costs.”).

³³⁹ *Id.* at 300, 301.

³⁴⁰ See, e.g., *Calles v. Scripto-Tokai Corp.*, 864 N.E.2d 249, 255, 256–57 (Ill. 2007) (stating that Illinois applies the consumer expectations test and the risk-utility test (citing *Blue v.*

impact on crash-optimization algorithms.

The Restatement (Second) of Torts provides that a product is defective if it is “dangerous to an extent beyond that which would be contemplated by the ordinary consumer who purchases it.”³⁴¹ Thus, this test does not consider the subjective expectations of an individual consumer; instead, it focuses on what a reasonable consumer would expect from a product.³⁴² As such, if a reasonable consumer would commit self-sacrifice in the Bridge Problem, the beliefs of the individual operator sacrificed to save the school bus full of children are irrelevant, and the product is not defective. One problem for car manufacturers under the consumer-expectations test is that over-hype about the abilities of autonomous vehicle to travel safely may create unreasonable expectations by society: members of the jury may expect autonomous vehicles to be accident free, or at least able to perform better than is technologically feasible.³⁴³ It is unclear, however, whether the consumer-expectations test would even be used in the states that have retained it. Some states limit the application of the consumer-expectations test to non-complex products.³⁴⁴ The California Supreme Court has stated: “[T]he consumer expectations test is reserved for cases in which the everyday experience of the product’s users permits a conclusion that the product’s design violated minimum safety assumptions, and is thus defective regardless of expert opinion about the merits of the design.”³⁴⁵ Because of the complexity involved in the operation of autonomous vehicles—especially the complexity of the crash-optimization algorithm—some scholars question whether the consumer expectations test could be used for autonomous vehicles.³⁴⁶

Because of the limitations of the consumer-expectations test, most

Envtl. Eng’g, Inc., 828 N.E.2d 1128, 1138–39 (2005)); *Delaney v. Deere & Co.*, 999 P.2d 930, 944 (Kan. 2000) (“Kansas . . . adheres to the consumer expectations test.”).

³⁴¹ RESTATEMENT (SECOND) OF TORTS § 402A cmt. i (AM. LAW INST. 1965).

³⁴² See, e.g., *Vincer v. Esther Williams All-Aluminum Swimming Pool Co.*, 230 N.W.2d 794, 798–99 (Wis. 1975) (“This is an objective test and is not dependent upon the knowledge of the particular injured consumer, although his knowledge may be evidence of contributory negligence under the circumstances.” (citing *Dippel v. Sciano*, 155 N.W.2d 55, 64–65 (Wis. 1967))).

³⁴³ See *Marchant & Lindor*, *supra* note 275, at 1334; *Garza*, *supra* note 18, at 600–01.

³⁴⁴ See, e.g., *Garza*, *supra* note 18, at 591 (“The more complex a product is, the more difficult it is to apply the consumer-expectations test.”).

³⁴⁵ *Soule v. Gen. Motors Corp.*, 882 P.2d 298, 308 (Cal. 1994) (emphasis omitted). A California Court of Appeals determined that an air bag is “not part of the ‘everyday experience’ of the consuming public.” *Pruitt v. Gen. Motors Corp.*, 86 Cal. Rptr. 2d 4, 6 (Cal. Ct. App. 1999).

³⁴⁶ See *Marchant & Lindor*, *supra* note 275, at 1324 (“[The consumer-expectations test] is generally considered particularly inapplicable in cases involving the analysis of technical and scientific information.”); *Garza*, *supra* note 18, at 600–01.

design defect claims will be analyzed under the risk-utility test.³⁴⁷ Under this test, a design defect occurs when:

[T]he foreseeable risks of harm posed by the product could have been reduced or avoided by the adoption of a reasonable alternative design by the seller or other distributor, or a predecessor in the commercial chain of distribution, and the omission of the alternative design renders the product not reasonably safe.³⁴⁸

A manufacturer could be faced with a products liability suit from the operator of the vehicle, or from the third party harmed by the decision made by the crash-optimization algorithm.³⁴⁹ In such a case, the court will need to determine whether the crash-optimization algorithm could have been more safely written. Proving a safer crash-optimization algorithm is an expensive endeavor: it will require various experts.³⁵⁰

Also, a design defect claim could be brought against the automaker even if the autonomous vehicle did not cause the accident.³⁵¹ For example, revisiting the revised Bridge Problem introduced in the discussion on negligence where a vehicle full of college students is behind the school bus and tries to pass the bus because the car's driver fails to see the oncoming autonomous vehicle. Assume that the autonomous vehicle turns into the school bus, killing the children and the bus driver. It is clear that the autonomous vehicle was not negligent in causing the accident; the driver of the car full of college students was at fault for the accident. However, the estates of the children and the school bus driver may bring a products liability suit against the autonomous vehicle manufacturer for a design defect; they would assert that the crash-optimization algorithm was defective for turning the vehicle into the school bus full of children instead of staying the course to hit the car.³⁵²

³⁴⁷ See, e.g., Owen, *supra* note 338, at 301 (“[M]ost modern courts have abandoned consumer expectations as the predominant test for design defectiveness.”).

³⁴⁸ RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2(b) (AM. LAW INST. 1998).

³⁴⁹ See Gurney, *supra* note 275, at 257, 258, 263, 272.

³⁵⁰ Glancy, *supra* note 275, at 1646–47 (asserting that litigation will be complex and costly when a connected autonomous vehicle is involved in an accident); Vladeck, *supra* note 6, at 139 (“[F]inding and retaining experts to put on a [risk-utility] case might be a dauntingly expensive enterprise for an individual plaintiff to bear.”); see also Gurney, *supra* note 275, at 263–64 (stating that the costs of pursuing a design defect claim make it infeasible to hold the manufacturer liable for many accidents caused by an autonomous vehicle).

³⁵¹ See Funkhouser, *supra* note 275, at 454.

³⁵² Note that although hitting the school bus in the revised Bridge Problem may be a defective product, hitting the school bus in the original Bridge Problem is probably not a design defect. Although driving off the bridge would prevent far less “harm” than striking the school

The Trolley Problem also creates an interesting interplay with the risk-utility test. Assume that the vehicle is going to hit the five pedestrians because the human driver fainted, and that the autonomous technology recognizes that the human driver's hands have been off the steering wheel for over five seconds. Yet, the autonomous technology does not take control over the vehicle so as to limit the manufacturer's tort liability.³⁵³ In such a case, the estates of the five persons may assert that the autonomous vehicle was defectively designed by not taking control when it knew that the operator's hands came off the steering wheel.³⁵⁴ This argument would certainly be stronger if the autonomous technology could have prevented the accident without causing the death of another person; in such a case, the autonomous vehicle would likely be found defective.

When faced with a products liability suit, the manufacturer is not without defenses, and two particular defenses are relevant: (1) comparative fault, and (2) state of the art.³⁵⁵ Comparative fault is a defense whereby the defendant's liability is reduced because the plaintiff was also negligent.³⁵⁶ The state of the art defense has been defined as "the technological feasibility of alternative safer designs in existence at the time the product was originally manufactured."³⁵⁷ In utilizing a state of the art defense, the car manufacturer would assert that it was not possible for the manufacturer to program its algorithm to prevent the harm to the plaintiff.³⁵⁸

4. Punitive Damages

Under any theory of liability, the plaintiff will try to recover punitive damages by asserting that the defendant acted intentionally or recklessly.³⁵⁹ Punitive damages are damages awarded in excess of a plaintiff's actual damages that represent a jury's moral condemnation of the defendant's conduct.³⁶⁰ Because the purpose of

bus, it is unlikely that an autonomous vehicle would be defective for not killing its operator—it is far more likely that the operator's estate would sue the manufacturer for self-sacrifice.

³⁵³ See Funkhouser, *supra* note 275, at 454.

³⁵⁴ See *id.*; Gurney, *supra* note 275, at 271.

³⁵⁵ See Gurney, *supra* note 275, at 266–67.

³⁵⁶ *Id.* at 267.

³⁵⁷ Lenhardt v. Ford Motor Co., 683 P.2d 1097, 1099 (Wash. 1984) (en banc).

³⁵⁸ See *id.*

³⁵⁹ See *Exxon Shipping Co. v. Baker*, 554 U.S. 471, 493–94 (2008).

³⁶⁰ *Cooper Indus. v. Leatherman Tool Grp., Inc.*, 532 U.S. 424, 432 (2001) ("A jury's . . . imposition of punitive damages is an expression of its moral condemnation." (citing *Gertz v. Robert Welch Inc.*, 418 U.S. 323, 350 (1974))).

punitive damages is to deter or punish,³⁶¹ such damages are typically reserved for situations in which the defendant's conduct was "outrageous, gross[ly] negligen[t], willful, wanton, and reckless[ly] indiffer[ent to] the rights of others, or [was] even more deplorable."³⁶²

The purpose of the crash-optimization algorithm is to mitigate harm when an accident is inevitable.³⁶³ The nature of that decision is beneficial to all involved, and it does not evidence egregious conduct on behalf of the algorithm writer. Society has been debating moral questions for millennia,³⁶⁴ and a court should ensure that a jury does not punish a car manufacturer merely out of disagreement with the decision made. Therefore, courts should be hesitant to allow a plaintiff to pursue punitive damages against an automaker for harm caused by a crash-optimization algorithm. An example of when punitive damages could serve a useful function could be in the Shopping Cart example: if the autonomous vehicle killed the baby because it did not want to cause damage to the car. There, the autonomous vehicle was likely programmed with "reckless disregard for the rights of others." However, this issue is nuanced because punitive damages are not warranted every time the vehicle hits the baby carriage: for example, when the autonomous vehicle malfunctions and hits the baby carriage. As should be clear, a court should be reluctant to allow a plaintiff to pursue punitive damages.

5. Summary

After discussing the tort liability of the manufacturer, it is necessary to discuss the implications that follow from such an imposition of liability on the manufacturer. In the Car Problem, an interesting corollary from imposing tort liability on the car manufacturer is that the manufacturer would likely be incentivized to hit the vehicle that would cost the least to repair.³⁶⁵ As indicated, the autonomous vehicle caused the accident when its brakes failed.

³⁶¹ See, e.g., *Exxon Shipping Co.* 554 U.S. at 492 ("Punitive[] [damages] are aimed not at compensation but principally at retribution and deterring harmful conduct."); *Cooper Indus.*, 532 U.S. at 432 ("[Punitive damages are] intended to punish the defendant and to deter future wrongdoing.").

³⁶² *Exxon Shipping Co.*, 554 U.S. at 493 (internal quotation marks omitted); see also S.C. CODE ANN. § 15-32-520(D) (2012) ("Punitive damages may be awarded only if the plaintiff proves by clear and convincing evidence that his harm was the result of the defendant's willful, wanton, or reckless conduct.").

³⁶³ *Why Ethics Matters*, *supra* note 14, at 72.

³⁶⁴ Goodall, *Ethical Decision Making*, *supra* note 1, at 62.

³⁶⁵ See *Why Ethics Matters*, *supra* note 14, at 72 (noting that an autonomous vehicle could minimize lawsuits by hitting the safer or larger vehicle).

Assume that the operator will not suffer severe harm in either choice. It seems evident from a standpoint of liability that the autonomous vehicle manufacturer will want to program its vehicle to hit the Jeep and not the Mercedes. Because the Jeep is bigger, less overall damage will occur to the vehicle, and whatever damage that does occur will cost less to repair.³⁶⁶ Therefore, the car manufacturer would probably program the vehicle to strike the Jeep.³⁶⁷

The effect of imposing tort liability on the car manufacturer becomes even more evident when there are two similarly-sized vehicles but one is more expensive. In the Car Problem, assume that instead of the Jeep Liberty the vehicle is a 2015 Toyota Camry; the MSRP for the Toyota Camry is \$22,970.³⁶⁸ In this case, the vehicles are relatively the same size—they are both sedans. The only difference is that the Mercedes retails at over twice that of the Camry.³⁶⁹ Therefore, hitting the Camry would result in less liability for the car manufacturer. This problem becomes even more exacerbated when the choice involves a used vehicle. The corollary from this discussion is that autonomous vehicles may be programmed to “protect” the rich.³⁷⁰

In the Motorcycle Problem, the automaker would be incentivized to program the vehicle to strike the helmeted—or responsible—motorcyclist.³⁷¹ As discussed in the introduction to the Motorcycle Problem, the motorcyclist who is not wearing a helmet has a higher risk of suffering brain injury and dying from an accident than the motorcyclist who wears a helmet. Because more harm will result to the motorcyclist who is not wearing a helmet, the self-interested manufacturer would want to program its vehicle to hit the helmeted motorcyclist.

In the Bridge Problem, the automaker would not be at fault for the accident because the bus driver caused the accident and the autonomous vehicle was not negligent in bringing about the accident. But the question is whether the crash-optimization algorithm will be

³⁶⁶ See *id.*

³⁶⁷ See Goodall, *Ethical Decision Making*, *supra* note 1, at 62 (“If the objective was to try and minimize cost, automated vehicles would choose to collide with the less expensive of the two vehicles, given a choice.”).

³⁶⁸ TOYOTA, *2015 Camry*, <http://www.toyota.com/camry/2015#!/Welcome> (last visited Jan. 18, 2016).

³⁶⁹ See MERCEDES-BENZ, *supra* note 121; TOYOTA, *supra* note 368.

³⁷⁰ See Goodall, *Ethical Decision Making*, *supra* note 1, at 62 (“If the objective was to try and minimize cost, automated vehicles would choose to collide with the less expensive of the two vehicles, given a choice.”).

³⁷¹ See *Why Ethics Matters*, *supra* note 14, at 73 (“[I]t seems reasonable to program a good autonomous car to swerve into the motorcyclist with the helmet.”).

able to determine who is “at fault” for bringing about the accident—fault is necessarily a legal consideration. Most likely, it will be unable to determine who caused the accident; instead, it would likely make its decision without knowing who caused the accident. Therefore, in the Bridge Problem, the automaker is faced with the question of bearing liability for the death of potentially a school bus full of children and a bus driver, damage to a school bus and autonomous vehicle, and physical harm to the occupant of the autonomous vehicle; or, alternatively, the death of the autonomous vehicle operator and destruction of its vehicle. Based on minimizing its liability, the automaker would program its vehicle to drive off the bridge.

Unlike in the Bridge Problem where the choice for the automaker is the death of many people or its customer, in the Tunnel Problem, the automaker is faced with the choice of killing only one person or its customer. Based on the manufacturer’s own self-interest, it would probably not program its autonomous vehicle to drive into the side of the mountain. Few people would buy an autonomous vehicle from a manufacturer that killed its occupants whenever it risked killing someone else. A manufacturer may be comfortable with killing its operator to save a greater number of people, but that number is likely not one person.

In the Trolley Problem, no automaker would likely program its technology to take control of the vehicle to save five persons when it means killing one person. When the human driver is operating the autonomous vehicle, that driver is at fault for all the harm caused; when the autonomous vehicle takes control over the vehicle, it would be responsible for the harm. Moreover, it is unlikely that a manufacturer would program its vehicle with a crash-optimization algorithm that takes control of the vehicle when doing so will cause harm to anyone or damage to property.

C. Criminal Law

In addition to implicating tort law, the decisions made by a crash-optimization algorithm may violate criminal laws. Generally, the government must prove two elements to convict a person of a crime: (1) *mens rea*, and (2) *actus reus*.³⁷² The *mens rea* refers to the mental state that a defendant must have before she can be convicted of a

³⁷² See, e.g., Jeffrey K. Gurney, *Driving into the Unknown: Examining the Crossroads of Criminal Law and Autonomous Vehicles*, 5 WAKE FOREST J.L. & POL’Y 393, 407 (2015).

crime;³⁷³ four common types of mental states are purposely, knowingly, recklessly, and negligently.³⁷⁴ However, an evil thought is not criminal; a person must also engage in criminal conduct (the *actus reus*).³⁷⁵ The *actus reus* is established by a voluntary act or an omission when a legal duty to act is imposed.³⁷⁶ Although the general requirement is that a defendant must have criminal intent, many automobile crimes are strict liability offenses, meaning that they do not require proof of a mental state.³⁷⁷ These offenses are typically “rules of the road” offenses, such as speeding.³⁷⁸

An autonomous vehicle cannot be criminally responsible for its own violations of criminal law.³⁷⁹ Although there is a “body to . . . kick[],” there is “no soul to . . . damn[].”³⁸⁰ Therefore, when an autonomous vehicle violates a criminal law, the persons who could be responsible are those who use the vehicle (the operator) or who control the vehicle (the car manufacturer).³⁸¹ Traditionally, the operator of the vehicle has been the responsible party for criminal violations.³⁸² The government imposes responsibility on the traditional driver because she is the one that caused the vehicle to act;³⁸³ however, with an autonomous vehicle, the operator does not necessarily violate any criminal laws by engaging in other activities.³⁸⁴ Thus, any criminal responsibility will not be based on operator’s actions.³⁸⁵ As discussed

³⁷³ See, e.g., 1 CHARLES E. TORCIA, *WHARTON’S CRIMINAL LAW* § 27, at 164 (15th ed. 1993).

³⁷⁴ MODEL PENAL CODE § 2.02(2) (AM. LAW. INST. 2015).

³⁷⁵ 1 TORCIA, *supra* note 373, § 25, at 138; see also 1 LAFAYE, *supra* note 257, § 6.1, at 423 (“One basic premise of Anglo-American criminal law is that no crime can be committed by bad thoughts alone.”).

³⁷⁶ 1 TORCIA, *supra* note 373, § 25, at 139.

³⁷⁷ See Douma & Palodichuk, *supra* note 252, at 1159 (stating that many offenses are strict liability offenses); Gurney, *supra* note 372, at 407 (discussing strict liability offenses).

³⁷⁸ See Gurney, *supra* note 372, at 407.

³⁷⁹ See, e.g., *United States v. Athlone Indus., Inc.*, 746 F.2d 977, 979 (3d Cir. 1984) (“[R]obots cannot be sued.”); Vladeck, *supra* note 6, at 121 (“[T]hese machines, notwithstanding their sophistication, have no attribute of legal personhood.”).

³⁸⁰ This quote was originally said by the Lord Chancellor of England when he was discussing a convicted corporation. See John C. Coffee, Jr., “No Soul to Damn: No Body to Kick”: *An Unscandalized Inquiry into the Problem of Corporate Punishment*, 79 MICH. L. REV. 386, 386 (1981).

³⁸¹ Vladeck, *supra* note 6, at 120–21 (“Any human . . . that has a role in the development of the machine and helps map out its decision-making is potentially responsible for wrongful acts—negligent or intentional—committed by, or involving, the machine.”).

³⁸² Gurney, *supra* note 372, at 410.

³⁸³ See *id.* at 411.

³⁸⁴ One article has described the role of the human in an autonomous vehicle as “playing the role of potted plant.” Vladeck, *supra* note 6, at 125–26.

³⁸⁵ See Claire Cain Miller, *When Driverless Cars Break the Law*, N.Y. TIMES (May 13, 2014), http://www.nytimes.com/2014/05/14/upshot/when-driverless-cars-break-the-law.html?_r=0&abt=0002&abg=1

(“Criminal law is going to be looking for a guilty mind, a particular mental state—should this person have known better?” [Ryan] Calo said. ‘If you’re not driving

elsewhere, punishing the operator for violations of criminal law while the vehicle is in autonomous mode is antiquated and should be reformed—such punishment serves little to none of the purposes of punishment.³⁸⁶

Because the traditional driver is the one who violates the criminal law,³⁸⁷ the car manufacturer is currently not liable for criminal violations of its automobiles. The vehicle does nothing, except perhaps provide a means for the criminal violation to occur. However, with autonomous vehicles, the automaker is no longer passively putting its vehicles on the market that may be used to violate criminal or traffic rules. When a vehicle in autonomous mode violates a traffic law or criminal statutes, such violation is caused by the manufacturer's programming and not the operator;³⁸⁸ a violation could occur either by a malfunction in the algorithm, causing the violation or because the algorithm was programmed to violate the traffic law in that instance. An example of the latter of the two situations is when an autonomous vehicle swerves into the opposite lane to avoid hitting a child. As is obvious, the programming of a traffic violation is beneficial in that example. No traditional driver would be prosecuted for violating a traffic law to avoid injuring a child; the benefit of the child's life outweighs the benefit of traffic law compliance. If, for some reason, an officer ticketed the traditional driver, she could use the defense of necessity.

Necessity is a justification defense that is sometimes called "choice-of-evils."³⁸⁹ "When the pressure of circumstances presents one with a choice of evils, the law prefers that he avoid the greater evil by bringing about the lesser evil."³⁹⁰ Thus, "[n]ecessity is, essentially, a utilitarian defense."³⁹¹ To assert a necessity defense, a defendant must prove:

- (1) that he was faced with a choice of evils and chose the lesser evil; (2) that he acted to prevent imminent harm; (3) that he reasonably anticipated a causal relation between his conduct and the harm to be avoided; and (4) that there were no other legal alternatives to violating the law.³⁹²

the car, it's going to be difficult.").

³⁸⁶ Gurney, *supra* note 372, at 411.

³⁸⁷ *Id.* at 414.

³⁸⁸ *See id.* at 418–19.

³⁸⁹ 2 WAYNE R. LAFAVE, SUBSTANTIVE CRIMINAL LAW § 10.1, at 117 (2d ed. 2003).

³⁹⁰ *Id.* § 10.1, at 118.

³⁹¹ *United States v. Schoon*, 971 F.2d 193, 196 (9th Cir. 1991).

³⁹² *United States v. Aguilar*, 883 F.2d 662, 693 (9th Cir. 1989) (citing *United States v. Dorrell*, 758 F.2d 427, 430–31 (9th Cir. 1985)), *superseded by statute on other grounds as stated*

An important fact about the necessity defense is that the defendant does not deny that she committed the crime; instead, the defendant argues that violating the law was the right decision.³⁹³ Some jurisdictions limit the application of the necessity defense to situations where the defendant did not cause the underlying need for necessity.³⁹⁴

Another limitation is that the necessity defense is inapplicable when the choice of evils requires the intentional killing of another person.³⁹⁵ The famous English case of *Regina v. Dudley and Stephens*³⁹⁶ cogently represents this principal. In that case, four seaman were stranded at sea on a dinghy for approximately nineteen days without food, water, or hope of rescue.³⁹⁷ On the twentieth day, two of them killed one of the others so that the remaining three could survive.³⁹⁸ Five days later they were rescued.³⁹⁹ Upon returning home, the Queen prosecuted the two men and they were convicted of murder, even though the jury found that all four seaman would have died had the two not killed the one, and that they had no prospect for relief when the two killed him.⁴⁰⁰

Autonomous vehicles will cause people to die, and the crash-optimization algorithm may be programmed such that someone may die as a result of its programming. In those situations, the question is whether the car manufacturer could be prosecuted. In the Trolley

in *Khan v. Holder*, 584 F.3d 773, 783 (9th Cir. 2009).

³⁹³ See John T. Parry, *The Virtue of Necessity: Reshaping Culpability and the Rule of Law*, 36 HOUS. L. REV. 397, 405 (1999).

³⁹⁴ See, e.g., DEL. CODE ANN. tit. 11, § 463 (2015); N.Y. PENAL LAW § 35.05(2) (McKinney 2015). Other states and the Model Penal Code do not prohibit the use of a necessity defense in all situations where the defendant's negligence or recklessness created the need for the necessity. See, e.g., HAW. REV. STAT. § 703-302 (2015); ME. REV. STAT. tit. 17-A, §§ 103, 103-A, 103-B (2015); MODEL PENAL CODE § 3.02 (AM. LAW INST. 2015). These statutes limit the defendant's culpability to the state of mind that caused the need for necessity. For example, if *A* drives recklessly and thereby creates a situation where he must either stay in the roadway and run down *B* and *C* or go on the sidewalk and strike *D*, and he chooses the lesser harm and hence strike and kills *D*, he is guilty of the recklessness type of manslaughter of *D* (on account of his reckless conduct in creating the situation) but not, it would seem, for the intentional murder of *D*. See 2 LAFAVE, *supra* note 389, § 10.1, at 125.

³⁹⁵ See *People v. Maher*, 594 N.E.2d 915, 916–17 (N.Y. 1992) (demonstrating that at least one court has stated that a jury can consider whether the underlying act by a driver—speeding—was out of necessity in defense to a criminal negligence prosecution); John Alan Cohan, *Homicide by Necessity*, 10 CHAP. L. REV. 119, 120 (2006) (“At common law and almost universally in modern law, the necessity defense has been consistently denied in cases where the actor commits intentional homicide in order to avert a greater evil.”).

³⁹⁶ *R v. Dudley and Stevens*, [1884] 14 QBD 273 (Eng.).

³⁹⁷ See *id.* at 273, 275.

³⁹⁸ *Id.* at 274.

³⁹⁹ *Id.*

⁴⁰⁰ *Id.* at 273, 275, 288.

Problem, the human driver is either intentionally or unintentionally going to kill five persons. To save those five persons, the autonomous technology takes control over the vehicle to avert the accident; in doing so, however, one person was killed. An argument could be made that the autonomous technology “intentionally” killed that one person: the technology took control knowing by doing so the person would die.⁴⁰¹ As can be seen by the discussion of *Dudley and Stephens*, necessity is not a defense to the death of the one person. However, prosecuting the car manufacturer for the deaths caused by its cars is impractical and infeasible.

The practical problem is which employee would be criminally liable for the so-called “murder” of the pedestrian. A piece of paper cannot be imprisoned. The CEO of a major car company like Ford would not write the algorithm. Presumably someone in the company made the decision—or actually, she may not have thought of the exact situation that caused the death. Assume that the vehicle is programmed to minimize harm; in such a case, the manager who made that decision did so with good intentions: to minimize harm. She would have no knowledge of the specific incident that gave rise to the person’s death because the decisions are so attenuated in time and space.⁴⁰² Although the company may know that its mitigation algorithm could result in someone’s death, that result was never its intent. Therefore, courts should not impose criminal liability on a manufacturer for the decisions made by the crash-optimization algorithm.

The issue of traffic tickets being issued to the manufacturer is equally troublesome. An autonomous vehicle will likely only violate traffic laws when it is malfunctioning, or to avoid a greater harm through use of its accident-avoidance or crash-optimization algorithms.⁴⁰³ When the vehicle malfunctions and violates a traffic

⁴⁰¹ See *Why Ethics Matters*, *supra* note 14, at 75 (“Again, in a real-world accident today, a human driver usually has neither the time nor the information needed to make the most ethical or least harmful decisions. A person who is startled by a small animal on an otherwise uneventful drive may very well react poorly. He might drive into oncoming traffic and kill a family, or oversteer into a ditch and to his own death. Neither of these results, however, is likely to lead to criminal prosecution by themselves, since there was no forethought, malice, negligence, or bad intent in making a forced, split-second reaction. But the programmer and OEM do not operate under the sanctuary of reasonable instincts; they make potentially life-and-death decisions under no truly urgent time-constraint and therefore incur the responsibility of making better decisions than human drivers reacting reflexively in surprise situations.”).

⁴⁰² See Goodall, *Machine Ethics*, *supra* note 1, at 60 (“[T]he decision itself is a result of logic developed and coded months or years ago.”).

⁴⁰³ See FLA. STAT. ANN. § 319.145(1)(d) (West 2015) (“An autonomous vehicle . . . must . . . [b]e capable of being operated in compliance with the applicable traffic and motor vehicle laws of this state.”); Goodrich, *supra* note 18, at 281 (“Presumably, every autonomous vehicle will be

law, the manufacturer will more than likely be punished—whoever is harmed will sue the company. When the vehicle violates the traffic law intentionally to prevent an accident, society should not punish the manufacturer. Therefore, a manufacturer should not be criminally liable when its vehicles violate traffic laws.

VI. COMBINING ETHICS AND LAW TO DEVELOP A LEGAL SYSTEM FOR CRASH-OPTIMIZATION ALGORITHMS

Parts I and II introduced autonomous vehicles. Part III introduced seven moral dilemmas. Part IV discussed utilitarian and Kantian ethics and their application to crash-optimization algorithms. Part V examined the application of tort and criminal law to crash-optimization algorithms. This Part ties the previous five Parts together to propose how the government should impose responsibility in these situations.

Initially, however, this Part examines whether these decisions should even be made. As indicated in Parts IV and V, the decisions create complicated and ambiguous moral questions and legal liability. Thus, an important question is whether society even wants autonomous technology to make these difficult moral decisions. After answering yes to that question, this Part then considers who should decide how the crash-optimization algorithm is programmed. Three different groups—or a combination thereof—could make the decision: (1) the autonomous vehicle owner; (2) the algorithm writer; or (3) the government. This Part concludes by developing a theory as to how these decisions should be treated legally.

A. *Does Society Want Robotic Cars Making These Decisions?*

A crash-optimization algorithm will decide deeply held moral questions, and it will implicate various criminal and tort law doctrines.⁴⁰⁴ In addition, the algorithm will greatly impact third parties.⁴⁰⁵ Thus, an initial consideration is whether society even wants an autonomous vehicle to make these decisions. “One way of dealing with [ethical dilemmas involving autonomous vehicles] is to avoid them altogether.”⁴⁰⁶ There are three ways to avoid having an

programmed to follow the rules of the road and will not violate traffic laws.”).

⁴⁰⁴ See Goodrich, *supra* note 18, at 280, 281; Millar, *supra* note 15.

⁴⁰⁵ See Goodall, *Ethical Decision Making*, *supra* note 4, at 95.

⁴⁰⁶ *Robot Ethics: Morals and the Machine*, ECONOMIST (June 2, 2012), <http://www.economist.com/node/21556234>.

autonomous vehicle make these ethical decisions: (1) ban autonomous vehicles; (2) require full attention of a human driver; or (3) make vehicles randomly decide whom or what to hit.

1. Ban Autonomous Vehicles

One solution to the problems discussed in this article is to ban vehicles that operate without human input. This would make it so a robot car never decides these ethical questions. Autonomous vehicles will cause accidents.⁴⁰⁷ They will kill people—and the people killed may not be the same people who have died in a non-autonomous vehicle marketplace.⁴⁰⁸

But many other technologies introduce these same risks and concerns.⁴⁰⁹ Vaccinations kill people and cause serious medical complications.⁴¹⁰ However, “[v]accines are one of medicine’s greatest accomplishments,”⁴¹¹ and are an example of a mass-produced good where the benefits are so great that society “should encourage vigorously.”⁴¹² Indeed, automobiles are themselves a risky endeavor. By permitting automobile use, society has determined that the benefits of such vehicles outweigh approximately 33,000 deaths, 2.3 million accidents, and all the other negative side effects that occur annually.⁴¹³

As indicated in Part II, autonomous vehicles are expected to save thousands of lives, billions of dollars in damages, and increase overall happiness. Therefore, as a whole, these vehicles will be immensely beneficial to society when implemented safely. Autonomous vehicles have the potential to provide the major benefit of traditional automobiles—transportation—with other benefits that a vehicle cannot currently provide, all while lowering the overall risk.⁴¹⁴ With those benefits, no one could seriously contend that society should ban

⁴⁰⁷ See, e.g., Marchant & Lindor, *supra* note 275, at 1321 (“Cars crash. So too will autonomous vehicles, a new generation of vehicles under development that are capable of operating on roadways without direct human control.”).

⁴⁰⁸ See *The Ethics of Saving Lives*, *supra* note 138.

⁴⁰⁹ See, e.g., Peter Huber, *Safety and the Second Best: The Hazards of Public Risk Management in the Courts*, 85 COLUM. L. REV. 277, 291 (1985) (“Everything is risky, and risk is everywhere. The natural state of the world is not safety but abundant (though often natural) hazard.”).

⁴¹⁰ *Id.* at 285 (“[V]accination, like everything else, is not perfectly safe. It can have extremely serious side effects, and occasionally even causes the disease it is supposed to prevent.”).

⁴¹¹ Eva B. Stensvad, Note, *Immunity for Vaccine Manufacturers: The Vaccine Act and Preemption of Design Defect Claims*, 95 MINN. L. REV. 315, 316 (2010).

⁴¹² See Huber, *supra* note 409, at 290.

⁴¹³ See *supra* Section II.B (discussing accidents and deaths caused by automobiles).

⁴¹⁴ See *supra* Section II.B.

autonomous vehicles.

Accordingly, the real discussion is not whether society should allow autonomous vehicles, but rather whether they will be safe enough *when* automakers introduce them into the marketplace.⁴¹⁵ Because the discussion is about *when* to allow these cars and not *if* they should be allowed, it is clear that the automaker will still need to address how to program a crash-optimization algorithm.

2. Require Full Attention

Another potential solution to the problem of how to program a crash-optimization algorithm is to require the autonomous vehicle operator to always pay attention to the vehicle. According to this theory, such attentiveness would allow the operator to intervene and make the difficult ethical decision.⁴¹⁶ This “solution” already has a template in autopilot for ships or airplanes.⁴¹⁷ Pilots are trained—and are expected—to pay attention while autopilot is flying the plane.⁴¹⁸ Likewise, operators could be trained—and be expected—to take control of the autonomous vehicle when the vehicle encounters a situation involving an unavoidable accident.

Initially, this system would defeat two major purposes of autonomous vehicles.⁴¹⁹ As indicated in Part II, two major purposes of these vehicles are to allow people to engage in other activities while

⁴¹⁵ Marchant & Lindor, *supra* note 275, at 1330. However, it is unlikely that an automaker would put an autonomous vehicle on the market unless it was safer than a human driver; otherwise, the manufacturer would “expose itself to lawsuits and runaway liability.” *Id.*

⁴¹⁶ Goodall, Presentation, *supra* note 50, at 1 (“Several automakers insist that the human driver will ultimately be responsible for monitoring the vehicle and the roadway, and in the event of a crash, the driver will be available to take control with little notice.”); *see also* WALLACH & ALLEN, *supra* note 12, at 15 (“[E]ngineers often think that if a (ro)bot encounters a difficult [ethical] situation, it should just stop and wait for a human to resolve the problem.”).

⁴¹⁷ *See, e.g.*, Marchant & Lindor, *supra* note 275, at 1325 (“[A]irplanes capable of flying on ‘autopilot’ (while also manned by a live pilot) provide a close analogy to autonomous vehicles.”); Colonna, *supra* note 26, at 93–99 (discussing the use of autopilot in airplanes and ships); LeValley, *supra* note 275, at 9 (“A natural analogy to autonomous vehicles is autopilots used in airplanes and ships.”).

⁴¹⁸ LeValley, *supra* note 275, at 9–10 (“In the contexts of airplanes and ships, constant oversight is both implied and expected, thus reducing the role of the autonomous technology.”); K. Krasnow Waterman & Mathew T. Henshon, *Imagine the Ram-If-Ications*, SCI. TECH LAW., Spring 2009, at 14, 15 (2009) (“[C]onstant human oversight is both implied and expected, to determine whether then-current use of the autopilot is appropriate.”); *see also* Bilger, *supra* note 16, at 106 (“[P]ilots are trained to stay alert and take over in case the computer fails.”). One court has stated that “[t]he obligation of those in charge of a plane under robot control to keep a proper and constant lookout is unavoidable.” *Brouse v. United States*, 83 F. Supp. 373, 374 (N.D. Ohio 1949).

⁴¹⁹ *See* Vladeck, *supra* note 6, at 121 n.16 (stating that requiring the driver to remain vigilant would defeat the goal of increased productivity).

commuting and to provide a means of transportation for those who cannot currently drive due to their age or disability.⁴²⁰ Requiring full attention would ban people from engaging in other activities and no disabled person or minor could legally comply with such a requirement. In addition to these problems, this “solution” is shortsighted. The ultimate goal is for autonomous vehicles to operate without human input—or for that matter, even without a human in the vehicle.⁴²¹ This theory would defeat that goal, too.

It is also unlikely that such an attentiveness requirement could resolve these ethical dilemmas. There is no indication that an operator could determine which decision to make in any given situation; even assuming the operator could decide the “right” decision, she probably could not safely intervene in time to execute that decision. For an operator to make the decision, that person needs “to be aware of what is taking place in the road, how surrounding traffic is behaving, and whether the automated vehicle is responding effectively.”⁴²²

The operator would need to pay attention at all times and be constantly thinking about how best to resolve the dilemmas the vehicle encounters. It is unreasonable and unrealistic to think that operators would even pay attention to the road.⁴²³ Current law requires drivers of traditional vehicles to pay attention at all times⁴²⁴—and the nature of current vehicles requires it too. Yet, drivers still distract themselves and impair their driving abilities.⁴²⁵ As stated in Part II, over ninety percent of accidents are caused by driver error.⁴²⁶ Thus, it is unrealistic to think that people will pay more attention in a vehicle that can operate safely without them

⁴²⁰ See *supra* text accompanying notes 60–63.

⁴²¹ See Goodrich, *supra* note 18, at 285–86 (noting that autopilot technology that relies on a pilot to be present is contrary to the ultimate goal of an autonomous vehicle). The Nevada Regulations even presume that an autonomous vehicle may be operated without a person actually present in the vehicle. NEV. ADMIN. CODE § 482A.020 (2015).

⁴²² Natasha Merat et al., *Highly Automated Driving, Secondary Task Performance, and Driver State*, 54 HUM. FACTORS 762, 763 (2012).

⁴²³ Goodall, Presentation, *supra* note 50, at 1 (“[I]t is unrealistic to assume that a human passenger can effectively monitor an automated vehicle[]”. On test tracks, passengers in automated vehicles spend more time looking away from the road than those in traditional vehicles.”).

⁴²⁴ See, e.g., GOVERNORS HIGHWAY SAFETY ASS’N, DISTRACTED DRIVING: SURVEY OF THE STATES 12 (2013) [hereinafter DISTRACTED DRIVING SURVEY] (noting that forty-seven U.S. States and the District of Columbia have specific laws prohibiting drivers from engaging in distracting secondary activities, such as cell phone use, while driving).

⁴²⁵ See, e.g., Bilger, *supra* note 16, at 107 (“[F]ully engaged drivers are already becoming a thing of the past. More than half of all eighteen-to-twenty-four-year-olds admit to texting while driving, and more than eighty per cent [sic] drive while on the phone.”).

⁴²⁶ See *supra* note 50 and accompanying text.

paying attention than they do in a vehicle that cannot safely drive without them. Studies have confirmed that semi-autonomous technology in vehicles leads people to engage in other activities.⁴²⁷ Realizing that people could not be trusted to intervene, Google changed its autonomous vehicle design and now intends to design its vehicle without a steering wheel, accelerator, or brake pedal.⁴²⁸

Even if humans were capable of paying attention, that ability does not mean the operator is capable of making the “right” decision—as is evidenced by the need for the “emergency doctrine” in tort law.⁴²⁹ The operator may be unable to sense all the options that are actually available to her, or she could just have a radical ethical view.⁴³⁰ Another problem related to human perception is that the operator may not be able to accurately calculate when her intervention is actually needed. An operator may overestimate the risk of harm and take control when a truly unavoidable accident is not about to occur.⁴³¹ Autonomous vehicles should be able to avoid accidents that a driver cannot currently avoid: the vehicle will be able to detect and calculate alternative routes more efficiently than a human. Thus, a human driver may perceive an unavoidable accident, while the autonomous vehicle was preparing to use its accident-avoidance algorithm to prevent any accident.⁴³² Therefore, an overzealous operator may cause more accidents by paying attention than if that operator had been distracted.

The problem of inattentiveness or overzealousness is not resolved by requiring the autonomous vehicle to notify the operator when she

⁴²⁷ See Goodall, *Machine Ethics*, *supra* note 4, at 96. See generally ROBERT E. LLANERAS ET AL., HUMAN FACTORS ISSUES ASSOCIATED WITH LIMITED ABILITY AUTONOMOUS DRIVING SYSTEMS: DRIVER’S ALLOCATION OF VISUAL ATTENTION TO THE FORWARD ROADWAY 92, 95 (2013); Natasha Merat et al., *supra* note 422, at 770 (indicating that subjects engaged in secondary tasks more frequently when driving semi-autonomous vehicles than while driving manually).

⁴²⁸ See *supra* note 29 and accompanying text (discussing Google’s new car design).

⁴²⁹ *Why Ethics Matters*, *supra* note 14, at 75 (“[A] human driver usually has neither the time nor the information needed to make the most ethical or least harmful decisions.”); see also *supra* Section V.B.2 (discussing the need for the emergency tort doctrine by defendants who have made an unreasonable decision when faced with an emergency).

⁴³⁰ See *Why Ethics Matters*, *supra* note 14, at 70, 71, 75 (explaining that a driver forced to choose between killing an elderly woman versus a child in the event that the driver must swerve her vehicle makes an unethical choice no matter which option is chosen, according to modern ethical standards, because the driver’s decision is based on the person’s age).

⁴³¹ See Smith, *supra* note 81, at 594 (“Drivers who speed around blind corners but fear traveling over bridges demonstrate this tendency to underestimate some risks and overestimate others.”); Vladeck, *supra* note 6, at 121 n.16 (stating that an operator may perceive accidents that the autonomous vehicle could avoid).

⁴³² See Vladeck, *supra* note 6, at 121 n.16.

needs to make an ethical decision.⁴³³ Studies have shown that “humans may require substantial warning time” to be able to assess the situation and drive the vehicle safely.⁴³⁴ Thus, if the operator was not paying attention to the road when she was notified, she would probably not be able to react in time to safely maneuver the vehicle.⁴³⁵ Therefore, the government should not try to resolve this problem by requiring the operator to pay attention to the road.

3. Should Autonomous Vehicles be Programmed According to Ethics?

Even though society will want autonomous vehicles to be programmed with a crash-optimization algorithm, it does not necessarily follow that the algorithm must be programmed according to an ethical theory. Instead, the vehicle could be programmed to make its “decisions through a random-number generator,”⁴³⁶ or by flipping a coin.⁴³⁷ For example, in the Motorcycle Problem, an odd number (or heads) means the helmeted motorcyclist is hit and an even number (or tails) means the unhelmeted motorcyclist is hit.

Although this theory seems strange, it has the benefit of not incentivizing people to engage in unfavorable behavior.⁴³⁸ Historically, people have had irrational fears about new technology or even felt threatened by it—luddites—and it is unlikely that autonomous vehicles will be treated any differently.⁴³⁹ Society wants people to wear helmets when on a motorcycle; we want people to drive safer vehicles.⁴⁴⁰ However, if a person knows that by wearing a

⁴³³ See Goodall, Presentation, *supra* note 50, at 1 (“[A]n automated vehicle could attempt to predict dangerous situations and alert the driver in advance.”).

⁴³⁴ *Id.*

⁴³⁵ *Why Ethics Matters*, *supra* note 14, at 72 (“[T]here may not be enough time to hand control back to the driver.”); see also Goodall, *Machine Ethics*, *supra* note 4, at 5 (“In an emergency, a driver may be unable to assess the situation and make an ethical decision within the available time frame. In these situations, the automated vehicle would maintain control of the vehicle, and by default be responsible for ethical decision making.”); Goodall, Presentation, *supra* note 50, at 2 (“In many potentially dangerous situations, the computer must take evasive action without waiting on the approval of its human driver.”);

⁴³⁶ *The Robot Car of Tomorrow*, *supra* note 6.

⁴³⁷ See John M. Taurek, *Should the Numbers Count?*, 6 PHIL. & PUB. AFF. 293, 303, 306 (1977) (suggesting that trade-off situations should be resolved by flipping a coin).

⁴³⁸ *The Robot Car of Tomorrow*, *supra* note 6.

⁴³⁹ Merriam-Webster defines a “luddite” as “one of a group of early 19th century English workmen destroying laborsaving machinery as a protest; *broadly*: one who is opposed to esp. technological change.” *Luddite*, MERRIAM-WEBSTER’S COLLEGIATE DICTIONARY (11th ed. 2003).

⁴⁴⁰ See, e.g., ADVOCATES FOR HIGHWAY & AUTO SAFETY, 2015 ROADMAP OF STATE HIGHWAY SAFETY LAWS: LETHAL LOOPHOLES, 15 (2015) (“[B]y an overwhelming majority (80%), Americans favor state laws requiring all motorcyclists to wear helmets.”).

helmet or by driving a safer vehicle she may be targeted by an autonomous vehicle, she may be incentivized not to wear a helmet or purchase the safest vehicle. The random-number generator has the benefit of removing this incentive to change behavior. With a random-number generator, the motorcyclist would always be better off wearing a helmet and the car owner will always be better off with the safest vehicle.

Although a random-number generator has the inherent benefit of not impacting human behavior, this method of optimizing crashes is obviously flawed. Initially, it seems unlikely that many people will change their decisions out of fear that an autonomous vehicle may target them. When autonomous vehicles enter the marketplace, the majority of the accidents caused on roadways will still be caused by human-driven vehicles. In those instances, the person will be better off with the safest vehicle, or with a helmet. In addition, driving an unsafe vehicle or riding a motorcycle helmetless is not a fail-safe plan to avoid being targeted: such person loses her “benefit” when others drive less safe vehicles or do not wear helmets. When that happens, those people who chose less safe vehicles or ride motorcycles without helmets will be in worse positions than they would have otherwise been.

Another flaw with using a random-number generator is that society has the benefit to plan these decisions. “People expect and tolerate human moral failures. However, they might not tolerate such failures in their machines.”⁴⁴¹ In the Shopping Cart Problem, society will not tolerate the baby carriage being hit instead of the shopping cart because the autonomous vehicle unfortunately drew the baby’s “number.” The baby’s estate will have a strong argument that the autonomous vehicle was defectively designed or that the manufacturer was negligent in designing its vehicle to use a random-number generator. Thus, this system would likely subject the manufacturer to potentially runaway liability.⁴⁴² Therefore, a crash-optimization algorithm should not make its decisions based on a random-number generator or a coin flip, and it should, instead, be programmed based on ethics.⁴⁴³

⁴⁴¹ WALLACH & ALLEN, *supra* note 12, at 71.

⁴⁴² See *Why Ethics Matters*, *supra* note 14, at 82 (“If ethics is ignored and the robotic car behaves badly, a powerful case could be made that auto manufacturers were negligent in the design of their product, and that opens them up to tremendous legal liability, should such an event happen.”).

⁴⁴³ See *id.* (“[R]obot cars will . . . need to have crash-optimization strategies that are thoughtful about ethics.”).

B. Who Should Decide the Decision an Autonomous Vehicle Makes in a Truly Unavoidable Accident?

Because autonomous vehicles will need to make difficult ethical decisions, a question arises about who should make that decision. Three parties could make this decision: (1) the car owner; (2) the car manufacturer; or (3) the government. In the Open Roboethics Initiative survey discussed in the introduction to the Tunnel Problem,⁴⁴⁴ the readers were also asked who should make the decision to kill the child or kill the operator. Forty-four percent believed that the operator should make the decision, thirty-three percent responded that the government should decide, twelve percent thought that the manufacturer make the decision, while eleven percent believed someone else should decide.⁴⁴⁵

The survey examined the reasoning behind this question, too. Of those who believe the passenger should make the decision, fifty-five percent responded that the passenger should make the life or death decision, fourteen percent expressed distrust toward the technology, twelve percent said they are the ones most affected, and ten percent responded it is part of their ownership interest.⁴⁴⁶ Of those who believe the manufacturer should make the decision, sixty-three percent based their decision on the manufacturer's expertise and understanding of the technology, while thirteen percent expressed distrust towards government making the decision.⁴⁴⁷ Of those who believe the government should make the decision, thirty percent think that lawmakers can make a fair, impartial decision, twenty-seven percent said "that lawmakers provide the most democratically legitimate decision making," another twenty-seven percent based it on the need of a universal set of rules, and twelve percent believed that the decisions are legal issues that require lawmakers.⁴⁴⁸

1. The Car Owner

As indicated in the survey, the most popular choice was letting the owner of the vehicle decide these tough ethical decisions.⁴⁴⁹ This

⁴⁴⁴ See *supra* Section III.D.

⁴⁴⁵ *Who Should Die?*, *supra* note 130.

⁴⁴⁶ *My (Autonomous) Car*, *supra* note 133.

⁴⁴⁷ *Id.*

⁴⁴⁸ *Id.*

⁴⁴⁹ Goodall, Presentation, *supra* note 50, at 5 ("[Some] have argued that automated vehicles are moral proxies for their owners and should therefore reflect their owners' individual morals as closely as possible."); see also Millar, *supra* note 15 ("[T]here are certain deeply personal

theory would be based on the fact that some moral decisions have such grave implications that it only seems fair that the owner should make the decision.⁴⁵⁰ Because there is no right answer to these moral dilemmas and the decisions may result in death or serious injury for the occupant or other persons,⁴⁵¹ naturally it makes sense to think that the car owner should have input into the decision.⁴⁵² A model for this type of decision-making already exists. In medical ethics, the individual patient or those whom the patient trusts makes the decision.⁴⁵³ A doctor explains complex medical decisions to the patient, and then the patient makes the decision based on the information provided.⁴⁵⁴ The doctor respects the patient's decision, regardless of the implications.⁴⁵⁵

Although similar to the decisions facing a patient, stark distinctions exist between the medical and automotive fields. First, surgeries are intrusive on the person, and there are serious problems with doctors intruding into a person against that person's will.⁴⁵⁶ Although autonomous vehicles will make decisions that greatly impact people's lives, these decisions are not any different than other decisions that are made on behalf of people. For example, a person who flies on an airplane entrusts these ethical decisions to the pilot, or someone who rides a city bus entrusts these decisions to the bus driver. Neither the pilot nor the bus driver consults passengers to see how they would like the difficult ethical decisions to be decided. Second, a patient's decision has no direct effect on other persons. Although loved ones are impacted when a patient decides to forgo

moral questions that will arise with autonomous cars that ought to be answered by drivers.”).

⁴⁵⁰ See Joseph Brean, *Technical Issues with Robot Cars Just Engineering Problems. The Moral Quandaries are Harder to Fix*, NAT'L POST (Sept. 12, 2014), <http://news.nationalpost.com/news/technical-issues-with-robot-cars-just-engineering-problems-the-moral-quandaries-are-harder-to-fix> (“[A]utonomous cars have the capacity for great benefit, but they will have to be set with preferences for scenarios that are hard to imagine in advance, sometimes with a strong moral aspect— so strong that neither robots nor designers ought to be involved.”); Millar, *supra* note 158 (discussing a recent poll where participants were asked questions regarding who should make such morally compromising decisions, and seventy-seven percent of respondents believed that either the government or the users of the vehicles should make the call).

⁴⁵¹ See Millar, *supra* note 157.

⁴⁵² See *id.*

⁴⁵³ Millar, *supra* note 15.

⁴⁵⁴ See WORLD MEDICAL ASSOCIATION, MEDICAL ETHICS MANUAL 43 (2nd ed. 2009), http://www.wma.net/en/30publications/30ethicsmanual/pdf/ethics_manual_en.pdf; Millar, *supra* note 15.

⁴⁵⁵ Millar, *supra* note 15.

⁴⁵⁶ See Millar, *supra* note 15; Allan Schwartz, *Surgery, Depression, and Anxiety*, AM. ADDICTION CTRS. (Apr. 22, 2010), <https://www.mentalhelp.net/articles/surgery-depression-and-anxiety/>.

treatment, the patient's decision does not mean other patients will go without treatment. However, with autonomous vehicles, a person's decision will directly affect other people. The decision to program a vehicle to follow radical self-preservation means that a baby may die if hitting the baby carriage means less damage to the vehicle than other alternatives.

Nonetheless, the foregoing discussion only means that the decision facing the patient and the owner of an autonomous vehicle is different. But difference alone does not mean that the car owner should not make the decision, and this Part will analyze how the owner could have input into the decisions made by a crash-optimization algorithm. The owner could provide input in two different ways. First, an autonomous vehicle could be programmed to notify the operator that an accident is about to occur and what course of action the operator wants the vehicle to perform. Second, an autonomous vehicle owner may select the ethics of her car prior to the time when the vehicle encounters the situation. Because this article discussed the impracticability of allowing the operator to make an "in the moment" decision in Part IV.A, this Part focuses solely on the vehicle being programmed—either pre-programmed or through use of "ethical apps"⁴⁵⁷—according to the beliefs of the owner.

Under this theory, the car manufacturer or its dealership would need to interview each new car owner about how she wants her car to make its ethical decisions. To actually implement this system, the manufacturer or dealership would need to know more about the person than whether she is a utilitarian or a Kantian. Some people may have strong utilitarian or Kantian views, as well as some views that are not perfectly utilitarian or Kantian.⁴⁵⁸ For example, some professed utilitarians would flip the switch in the traditional Trolley Problem, but not want the surgeon to kill her patient for his organs.⁴⁵⁹ Instead, this method would need to be situation specific to learn about what the person actually believes. A prospective consumer could watch ethical scenarios and pick which choice she would make. For example, a consumer could watch the Tunnel

⁴⁵⁷ See *A Point of View: The Ethics of the Driverless Car*, BBC (Jan. 24, 2014), <http://www.bbc.com/news/magazine-25861214> ("[T]here should be a variety of ethical engines to install in your self-driving car."); *supra* Section IV.A.

⁴⁵⁸ See Wallach, *The Challenge of Moral Machines*, *supra* note 3, at 6 ("[M]ost people's moral intuitions do not conform to simple utilitarian calculations."); Wright, *supra* note 231, at 280.

⁴⁵⁹ See Wallach, *The Challenge of Moral Machines*, *supra* note 3, at 6 (stating that most people elect to flip the switch to save five persons but would not push the fat man off the bridge to save five persons). For a discussion of the trolley and surgeon problems, see Parts III.E and IV.B.

Problem from the viewpoint of (a) hit the child or (b) drive into the side of the mountain. That consumer could then select (a) or (b) to indicate which choice she would prefer. Based on her decision, the programmers would then be able to learn about this individual consumer's beliefs.

Assuming that this approach is technologically feasible, it has a promising benefit: it would alleviate fears of robot cars killing their occupants. A person would be less trusting of an autonomous vehicle when she has no input into its ethical decisions.⁴⁶⁰ If that person could make the choice herself, she may be more accepting of the technology.

However, this would also create ambiguous legal responsibility. As discussed in Part VI.B., manufacturers will likely be liable for accidents caused in autonomous mode. When the car owner provides input into the programming, such decision-making power may shield the manufacturer from liability in some, but not all, cases.⁴⁶¹ If the consumer chose to commit self-sacrifice in the event of the Tunnel Problem, it will be difficult for her estate to sue the car manufacturer for tort liability.⁴⁶² The car would not be defective if it was programmed in accordance with the owner's beliefs. Although it would shield the car manufacturer from liability in lawsuits brought by its consumer or her estate, the manufacturer may not be able to hide behind its consumer's decision in lawsuits brought by other people—including non-owner operators.⁴⁶³ For example, assume that, in the revised Bridge Problem, the consumer decides to hit the school bus instead of staying the course to hit the car or driving off the bridge. The estates of the children will argue that the consumer should not have been given the choice to hit the school bus, and that the consumer's pockets are not "big enough" to compensate all the children, making the manufacturer responsible for the harm.⁴⁶⁴ This decision-making power also has a negative effect on the owner of the

⁴⁶⁰ See Boeglin, *supra* note 275, at 178, 180. As some have acknowledged, many drivers may already be hesitant to cede control to a computer driver. *Id.* When they do not have say into these difficult ethical questions, they may be even more hesitant to cede control or buy an autonomous vehicle.

⁴⁶¹ But see Patrick Lin, *Here's a Terrible Idea: Robot Cars with Adjustable Ethic Settings*, WIRED (Aug. 18, 2014), <http://www.wired.com/2014/08/heres-a-terrible-idea-robot-cars-with-adjustable-ethics-settings/> (arguing that it would not shield the manufacturer from liability).

⁴⁶² See Colonna, *supra* 26, at 116 ("[A]utonomous car owners will inevitably blame their cars for crashes.").

⁴⁶³ See Lin, *supra* note 461.

⁴⁶⁴ Perhaps, the insurance industry would adjust premiums based on the decisions that the owner selected. Therefore, if the owner decides to strike the school bus, the owner would need to keep higher limits than a similar driver who elects self-sacrifice.

autonomous vehicle: the owner will no longer be able to hide behind the decision of the autonomous vehicle manufacturer. Therefore, it may not be in the owner's best financial interest to have a direct input into these decisions.

In any event, it is evident that the owner of the autonomous vehicle would not be able to make a full-fledged decision. Society will not want owners to make her programming decisions on arbitrary grounds, such as based on the person's race, gender, religious beliefs, sexual orientation.⁴⁶⁵ Indeed, the Professional Codes of Ethics of Engineers may not permit its members to program an autonomous vehicle to target people for certain reasons.⁴⁶⁶ The Institute of Electrical and Electronics Engineers (IEEE) requires its members "to treat fairly all persons and to not engage in acts of discrimination based on race, religion, gender, disability, age, national origin, sexual orientation, gender identity, or gender expression."⁴⁶⁷ Thus, an engineer who abided by the IEEE standards could not allow its consumers to take any of those factors into account.⁴⁶⁸ Therefore, the car manufacturer would need to make a decision concerning the relevant ethical criteria that an operator could choose from.

2. The Car Manufacturer

One troublesome aspect about allowing the car owner to decide how to mitigate liability is that the car owner may not be legally responsible for the accident. For example, when a malfunction in the autonomous technology causes the accident, most legal experts project that the car manufacturer will be legally responsible for the accident.⁴⁶⁹ Because the manufacturer is the party likely responsible for the underlying harm,⁴⁷⁰ it seems unfair to allow the car owner to determine how to limit the manufacturer's liability.

In addition, the manufacturer is in the best position to ensure that the vehicles are programmed according to ethics: it writes the algorithm and produces the vehicle.⁴⁷¹ Thus, the automaker is in the

⁴⁶⁵ Lin, *supra* note 461 ("Imagine that manufacturers created preference settings that allow us to save hybrid cars over gas-guzzling trucks, or insured cars over uninsured ones, or helmeted motorcyclists over unhelmeted ones. Or more troubling, ethics settings that allow us to save children over the elderly, or men over women, or rich people over the poor, or straight people over gay ones, or Christians over Muslims.").

⁴⁶⁶ *Why Ethics Matters*, *supra* note 14, at 70.

⁴⁶⁷ INST. OF ELECTRICAL & ELECTRONICS ENGINEERS, *supra* note 128.

⁴⁶⁸ See *Why Ethics Matters*, *supra* note 14, at 70.

⁴⁶⁹ See *supra* note 275 and accompanying text; Gurney, *supra* note 275, at 251 n.33, 258.

⁴⁷⁰ See Gurney, *supra* note 275, at 251 n.33, 258.

⁴⁷¹ See Anderson & Anderson, *supra* note 80, at 74, 75, 76 (assuming that the inventor makes

best technical position to program its vehicles with an ethical crash-optimization algorithm. Because of the inability to frame the relevant top-down rules and because of the difficulty that unique circumstances create for any system of rules, it is unlikely that every situation will be able to be adequately captured by a top-down crash-optimization algorithm.⁴⁷² One way to fill the gaps is through use of a bottom-up approach to ethics, which only the manufacturer can program.⁴⁷³

As indicated in Part IV, a bottom-up approach to ethics is a procedure whereby an autonomous vehicle would learn ethics “through the observation of human actions or through rewards for their own moral behavior.”⁴⁷⁴ An issue with watching how humans behave in unavoidable crash situations is that humans may not behave optimally—the decisions are “made quickly and under great stress, with little forethought or planning.”⁴⁷⁵ Thus, the car manufacturer would want to take great care to ensure that the vehicle is watching humans to determine what they believe rather than how they behave.⁴⁷⁶

Although the car manufacturer may be in the best position to make these decisions, the car manufacturer will face two constraints that may lead it to program the vehicles not necessarily according to ethical theory. The first constraint will come from the marketplace. The market would likely force car manufacturers to create selfish autonomous vehicles.⁴⁷⁷ Many people will be skeptical of autonomous vehicles, and the idea of an autonomous vehicle committing self-sacrifice will be frightening.⁴⁷⁸ Therefore, automakers will likely

the ethical decisions).

⁴⁷² See Anderson & Anderson, *supra* note 80, at 74 (“It would be extremely difficult, if not impossible, to anticipate every decision a robot might ever face and program it so that it will behave in the desired manner in each conceivable situation.”); Goodall, *Ethical Decision Making*, *supra* note 1, at 61–62 (“As vehicles continued to encounter novel situations, freak accidents, new road designs, and vehicle system upgrades would need to be addressed.”); Wallach, *The Challenge of Moral Machines*, *supra* note 3, at 6 (“For simple applications, the designers and engineers can anticipate all the situations the system will encounter, and can program in appropriate responses. However, some method to explicitly *evaluate* courses of action will need to be programmed into any (ro)bot likely to encounter circumstances the designers could not anticipate.”).

⁴⁷³ See WALLACH & ALLEN, *supra* note 12, at 112, 114; Goodall, *Ethical Decision Making*, *supra* note 1, at 62.

⁴⁷⁴ Goodall, *Ethical Decision Making*, *supra* note 1, at 62.

⁴⁷⁵ *Id.* at 60, 62 (noting that humans do not always behave optimally in these situations).

⁴⁷⁶ See *id.* at 62 (discussing how artificial intelligence algorithms should capture ideal behaviors).

⁴⁷⁷ See Goodall, Presentation, *supra* note 50, at 6 (indicating that programmers will likely make their vehicles self-protectionist).

⁴⁷⁸ See, e.g., Erik Sofge, *The Mathematics of Murder: Should a Robot Sacrifice Your Life to*

program their vehicles with a bias in favor of their consumers—even if it is not the best ethical decision. An automaker that does not create overly self-interested vehicles would face fierce advertisement from competitors, potentially leading to a competition over which automaker can make its vehicles the most protectionist.

The second constraint stems from imposing liability on the manufacturer. The manufacturer will try to minimize its liability.⁴⁷⁹ One implication of the manufacturer minimizing its liability is that the vehicle may not be programmed to prevent harm or damage caused when the operator is driving the vehicle. As discussed in Part V, the automaker would likely not program its technology to take control of the vehicle to prevent a greater harm when doing so will lead to a lesser harm. Failing to prevent an accident means no liability—causing an accident means liability.⁴⁸⁰ As also discussed in Part V, the manufacturer may not take into account important societal values—when doing so will result in more harm—or that the vehicles may be programmed with an inherent bias in favor of wealthy people.

3. The Government

In light of the limitations facing the car manufacturer, state governments or the federal government could enact a law outlining how the ethical decisions should be made—hereinafter, the “Safety Protocol.” The Safety Protocol could proscribe ethical rules for algorithm writers to use when programming their vehicles. It could also create a safe harbor for an autonomous vehicle manufacturer whereby compliance with the Safety Protocol results in some sort of immunity for the autonomous vehicle manufacturer.⁴⁸¹

This approach has many benefits.⁴⁸² First, it would provide

Save Two?, POPULAR SCI.: ZERO MOMENT (May 12, 2014), <http://www.popsci.com/blog-network/zero-moment/mathematics-murder-should-robot-sacrifice-your-life-save-two/> (“[N]obody in their right mind would buy an autonomous car that explicitly warns the customer that his or her safety is not its first priority.”).

⁴⁷⁹ See Jameson M. Wetmore, *Redefining Risks and Redistributing Responsibilities: Building Networks to Increase Automobile Safety*, 29 SCI., TECH., & HUM. VALUES (RECONSTRUCTING ORDER THROUGH RHETORICS OF RISK) 377, 385 (2004) (discussing the historic reluctance of automobile manufacturers to adopt new technologies out of fear of increased legal responsibility and liability).

⁴⁸⁰ See *Why Ethics Matters*, *supra* note 14, at 79.

⁴⁸¹ See *The Ethics of Saving Lives*, *supra* note 138; Steve Wu, *Panel Discussion at Association of Defense Counsel (ADC)*, STANFORD LAW SCH.: ROBOTICS & THE LAW (Feb. 19, 2011), <http://blogs.law.stanford.edu/robotics/2011/02/19/panel-discussion-at-association-of-defense-counsel-adc-dec-9-2010>.

⁴⁸² The author is cognizant of the fact that legislators would have difficulty agreeing out

certainty to automakers when programming their vehicles. Uncertainty is one of the greatest deterrents of new technology, and much uncertainty exists about how the law will treat autonomous vehicles.⁴⁸³ In addition to the general uncertainty facing autonomous vehicles, automakers will need to determine which ethical theory to use when programming their vehicles.⁴⁸⁴ Society has yet to agree on what ethical theory is “right.”⁴⁸⁵ This approach would provide a bright-line ethical theory for manufacturers to follow. Second, this approach would set expectations about the decisions made by a crash-optimization algorithm.⁴⁸⁶ Philosophers, engineers, psychologists, lawyers, and other relevant parties can provide beneficial input into the decisions made by a crash-optimization algorithm.⁴⁸⁷ Third, this approach would provide a sense of legitimacy to the eventual decisions made by the autonomous vehicles. When the decisions are made by computer scientists and engineers of a company, the decision has an air of secrecy to it. However, a decision seems more legitimate when the public has an input into the decisions through their elected officials.

Although the government’s intervention has those three benefits, the government may not be able to resolve all of the issues relevant to a crash-optimization algorithm. Instead, the government would be restricted to general principles with the manufacturer gap filling through bottom-up approaches to ethics. Such general rules should utilize the benefits of bright-line rules with the completeness of a utilitarian approach.

As to the bright-line rules, the government could prohibit the vehicles from being programmed to decide who to hit because of person’s race, gender, religion, political affiliation, sexual orientation, or any other arbitrary ground. The government could also issue general rules about actual accidents. For example, the government could establish that property damage is always preferred to injury to or death of a human and those injuries are more preferable than

what ethical rules should apply. See Hubbard, *supra* note 273, at 1871 (stating that objections to policy and political resistance hinder the government’s ability to pass laws).

⁴⁸³ See M. Ryan Calo, *Open Robotics*, 70 MD. L. REV. 571, 576 (2011) (noting that legal uncertainty has the potential to deter investments in robotic technology); Hubbard, *supra* note 273, at 1865, 66 (discussing liability law and the balancing act that takes place when introducing new technologies between the concern for physical safety and the desire for technological advancements).

⁴⁸⁴ Goodall, *Ethical Decision Making*, *supra* note 1, at 63.

⁴⁸⁵ See *Why Ethics Matters*, *supra* note 14, at 70 (providing an example of two ethical choices, but stating that neither choice is ethically correct).

⁴⁸⁶ *Id.* at 82.

⁴⁸⁷ See *id.*; Goodall, *Ethical Decision Making*, *supra* note 1, at 63.

death.⁴⁸⁸ Another rule could be that two deaths are worse than one death.⁴⁸⁹

However, such rules may not always lead to the best result or resolve all ethical situations. When deciding what rules to impose, the government will not have the benefit of using the 100% certainties that this article has used for the purpose of simplicity. The autonomous vehicle never, or very rarely, encounters these situations with 100% certainties; instead, it will face decisions based on risk that a result may occur.⁴⁹⁰ Certainly an autonomous vehicle should always choose to prevent severe harm to a person at the cost of damage to property, but the autonomous vehicle will rarely face a one-or-the-other choice. The autonomous vehicle will face risks of harm.⁴⁹¹ For example, an autonomous vehicle may be in a situation where an accident must occur: Choice A means that there is a 90 percent chance that \$400,000 in property damage occurs; or Choice B includes a .00001 percent chance that a person is severely injured. Which choice should the autonomous vehicle make in that case? An autonomous vehicle programmed to *always* protect people over property may choose Choice A—although that choice is not necessarily the best choice. Given the minute risk of harm that could occur to the person in Choice B, Choice A may lead to unnecessary destruction of property when no one would have likely been harmed.

Those proposed rules also do not necessarily provide a solution to all of the potential problems. For example, in the Bridge Problem, the autonomous vehicle may actually be faced with three decisions: (1) turn left off the bridge; (2) crash head on into the school bus; or (3) attempt to squeeze between the bus and the edge of the bridge on the right.⁴⁹² The first two options guarantee an accident; while the third option has a high probability of a severe crash occurring and a low probability of no crash.⁴⁹³ The “bright-line rules” stated above would not be able to decide what decision the vehicle should make because there is a high probability of death to people in all three cases. Thus, the autonomous vehicle would need a combination of general rules with a rule that can provide an answer to every situation; one such rule is that an autonomous vehicle should be

⁴⁸⁸ Goodall, *Ethical Decision Making*, *supra* note 1, at 63.

⁴⁸⁹ See *supra* Section IV.A (discussing utilitarianism).

⁴⁹⁰ See Goodall, Presentation, *supra* note 50, at 2 (stating that an advanced autonomous vehicle continuously calculates risk).

⁴⁹¹ See *id.* (discussing an example of a risk an autonomous vehicle may be faced with).

⁴⁹² Goodall, *Ethical Decision Making*, *supra* note 1, at 60.

⁴⁹³ *Id.* at 61.

programmed to minimize the total amount of expected harm that results from an accident. Under this rule, the autonomous vehicle would need to calculate the total expected amount of harm that would result from each choice, and the risk that each choice has of realizing that harm. This rule has the benefit of being complete: the autonomous vehicle will always be able to make a decision. Each decision is a matter of multiplying the risk of harm by the expected harm that would result from that choice, and comparing it to the other choices the autonomous vehicle could make. The vehicle would select the choice with the least amount of harm.

The minimize-harm rule could not be the only top-down rule because of its limitations. First, such rule fails to take into account important societal values. For example, in the Motorcycle Problem, a minimize-harm autonomous vehicle would hit the motorcyclist who wore a helmet solely because she wore a helmet. Although that result seems unfair, the minimize-harm algorithm would not be able to take that factor into account. Second, as indicated in Part III, the autonomous vehicle may kill a person instead of destroying property—of course, this depends on how the vehicle “values” human life. People could also argue that the minimize-harm rule has other problems. For instance, this algorithm could lead the autonomous vehicle to kill an innocent person when doing so would result in less harm or damage. For example, in the Trolley Problem, the autonomous technology would take control of the vehicle and swerve into the one pedestrian. In other situations, this rule may benefit the at-fault party. Modifying the Bridge Problem, assume that oncoming in the opposite lane of the autonomous vehicle is an SUV with one person in it. Behind the SUV is a vehicle full of intoxicated college students who are drinking while driving. The driver of the college student vehicle determines to pass the SUV; he does not realize that the autonomous vehicle is oncoming because he is drunk. In such a situation, the autonomous vehicle has the decision to (1) drive off the bridge, killing its occupant; (2) turning left to collide with the SUV, causing a serious collision; or (3) doing nothing and hitting the car head on. An autonomous vehicle programmed to minimize harm would not choose option three; certainly hitting the SUV with its one occupant would be better than hitting a vehicle full of people head on. Thus, the minimize-harm autonomous vehicle would “reward” the at-fault party and “punish” others. In light of those limitations to a crash-optimization algorithm that solely focused on minimizing harm, the autonomous vehicle will also need bright-line rules.

*C. Removing the Disincentives for the Manufacturer to Program
Ethical Cars*

Based on the foregoing, it is unlikely that any one party—the car owner, the manufacturer, or the government—can fully resolve the issue of crash-optimization algorithms.⁴⁹⁴ An autonomous vehicle programmed solely by the car manufacturer may be biased heavily in favor of the consumer, and focused on minimizing liability and not maximizing ethics. The government is unable to devise rules and create a system that applies in every situation. And the owner, if she is consulted specifically on some of the trickier questions like the Trolley Problem, is unable to decide all the ethical questions. Therefore, programming the crash-optimization algorithm will require collaboration amongst all parties—or at least the car manufacturer, and the government. Although determining exactly how the vehicle should be programmed is beyond the scope of this article, this subsection asserts that the government should, in acknowledgement of the difficulties facing the car manufacturer in these situations, clarify the automakers' legal responsibility.⁴⁹⁵

Legal uncertainty surrounding autonomous vehicles deters their introduction in the marketplace.⁴⁹⁶ Given the potential benefits that autonomous vehicles are likely to provide to society, government should facilitate their introduction into the marketplace. As has been seen throughout this article, the decisions made by crash-optimization algorithms implicate many unsettled ethical debates and create troublesome legal problems for operators and manufacturers of autonomous vehicles. The decision to program an autonomous vehicle to minimize damage may subject the autonomous vehicle manufacturer, operator, or both to tort and criminal liability.⁴⁹⁷ Therefore, the system, as it stands now, should

⁴⁹⁴ See Beiker, *supra* note 50, at 1153 (“Single organizations or institutions cannot solve the challenges of individual mobility. Nor can these challenges be resolved by reference to a single discipline. In the end, interested entities within industry, academia, and government need to work together, especially when addressing interdisciplinary topics in an emerging field.”).

⁴⁹⁵ See, e.g., Stephen P. Wood et al., *The Potential Regulatory Challenges of Increasingly Autonomous Motor Vehicles*, 52 SANTA CLARA L. REV. 1423, 1426 (2012) (“Government actions can influence the extent and the speed with which autonomous driving technologies are adopted.”).

⁴⁹⁶ See Calo, *supra* note 483, at 576 (“[L]egal uncertainty could discourage the flow of capital into robotics or otherwise narrow robot functionality, placing the United States behind other countries with a higher bar to litigation and a head start on research, development, and production.”).

⁴⁹⁷ See *id.* at 596–97 (“Some states even shift the burden of proof to the manufacturer-defendant to prove that the harm at issue was *not* foreseeable.”).

be clarified to provide certainty to autonomous vehicle manufacturers and future owners of the cars. This subsection recommends providing partial legal immunity to the autonomous vehicle manufacturers to incentivize them to take into account ethics in the programming of the autonomous vehicles.

The government has intervened to provide immunity and certainty to manufacturers or producers in certain industries.⁴⁹⁸ Congress enacted laws to limit liability for Y2K-related problems; the nuclear power industry; oil spill liability; vaccine manufacturers; and small plane and small plane part manufacturers.⁴⁹⁹ States have enacted numerous laws limiting liability, such as through “tort reform” or to limit punitive damages.⁵⁰⁰ Michigan’s autonomous vehicle statute provides immunity for car manufacturers and subcomponent system producers when another company has added its autonomous vehicle technology to their cars or their subcomponent systems.⁵⁰¹ Originally, Google was testing its autonomous technology on vehicles manufactured by other companies, such as Toyota Prii.⁵⁰² Thus, this law provided Toyota and other car manufacturers certainty that they would not be responsible for a car accident that resulted from Google’s modifications to their vehicles.⁵⁰³ The federal government or state governments should pass similar laws to provide certainty to autonomous vehicle manufacturers about their criminal and tort liability for accidents involving an autonomous vehicle.

⁴⁹⁸ See Marchant & Lindor, *supra* note 275, at 1337.

⁴⁹⁹ *Id.* at 1337–38.

⁵⁰⁰ See *id.* at 1338.

⁵⁰¹ MICH. COMP. LAWS § 600.2949b (2015). The statute reads:

(1) The manufacturer of a vehicle is not liable and shall be dismissed from any action for alleged damages resulting from any of the following unless the defect from which the damages resulted was present in the vehicle when it was manufactured:

(a) The conversion or attempted conversion of the vehicle into an automated motor vehicle by another person.

(b) The installation of equipment in the vehicle by another person to convert it into an automated motor vehicle.

(c) The modification by another person of equipment that was installed by the manufacturer in an automated motor vehicle specifically for using the vehicle in automatic mode.

(2) A subcomponent system producer recognized as described in section 244 of the Michigan vehicle code, 1949 PA 300, MCL 257.244, is not liable in a product liability action for damages resulting from the modification of equipment installed by the subcomponent system producer to convert a vehicle to an automated motor vehicle unless the defect from which the damages resulted was present in the equipment when it was installed by the subcomponent system producer.

Id.

⁵⁰² See Markoff, *supra* note 52.

⁵⁰³ See MICH. COMP. LAWS § 600.2949b.

1. Criminal Law

The autonomous vehicle manufacturer should receive complete criminal immunity from any suit resulting from the application of a crash-optimization algorithm and from use of its vehicles. As discussed in Part V, it seems unlikely the car manufacturer or algorithm writer could be criminally prosecuted when a crash-optimization algorithm “targets” someone to minimize harm, such as in the Trolley Problem. Moreover, as discussed in that Part, most criminal law violations committed by the vehicle in autonomous mode could be remedied through use of the tort system, and imposing criminal liability is infeasible and would deter the manufacture of these vehicles. Therefore, the government should provide certainty to manufacturers by providing immunity from criminal prosecution for decisions made by their autonomous vehicles.

The operator of the autonomous vehicle should not receive complete immunity from criminal punishment. Certainly, an operator should be criminally responsible for traffic violations when she is driving the vehicle. As discussed elsewhere, it makes little sense, however, to punish an operator for traffic violations that occur while the vehicle is in autonomous mode, or to criminally punish someone for harm caused by her autonomous vehicle.⁵⁰⁴ But, in some situations, for example, where an operator modifies her vehicle and it causes traffic violations or harm, or when she does not retake control over the vehicle when she knows that it is malfunctioning, criminal liability should be imposed on the operator.⁵⁰⁵

2. Tort Law

As to tort law, the immunity that government should afford to autonomous vehicles is more nuanced than that of criminal law. The automaker should not be civilly liable every time one of its vehicles is involved in an accident.⁵⁰⁶ Instead, the automaker should only be responsible for harm when its autonomous technology caused the underlying accident. Therefore, when an autonomous vehicle causes the situation that necessitates the use of the crash-optimization algorithm by an error, bug, or malfunction in its technology, the autonomous vehicle manufacturer should not be entitled to immunity

⁵⁰⁴ Gurney, *supra* note 372, at 417, 427.

⁵⁰⁵ *See id.* at 426.

⁵⁰⁶ *See* Hubbard, *supra* note 273, at 1865 (“Liability law is designed to achieve an efficient balance between the concern for physical safety and the desire for innovation.”).

and should be liable for the resultant harm. When that happens, the crash-optimization algorithm mitigates the manufacturer's liability,⁵⁰⁷ and thus, it is in the manufacturer's best interest to program its algorithm to minimize damage.

However, in situations where the autonomous vehicle did not cause the situation necessitating the use of the crash-optimization algorithm, the autonomous vehicle should be granted immunity so long as the decision made by the algorithm was not egregious. Although it is rudimentary tort law that someone is not liable when she is not at fault, without some formal protection,⁵⁰⁸ the automaker—because of the size of its pockets—will always be sued for accidents involving its cars.⁵⁰⁹ In addition, as indicated in Part V, tort law does not incentivize manufacturers to program their vehicles to prevent harm in all instances when they could, and immunity would remove the disincentive to minimize harm for automakers. Society will be better off if autonomous vehicles minimize all harm and not just the harm that they cause. Punishing a manufacturer when it minimizes other people's negligence—whether it is the operators', other drivers', or pedestrians' negligence—would be unfair, and it would create a disincentive for them to produce these vehicles and to prevent that harm.

The “black boxes” in autonomous vehicles could enable this system.⁵¹⁰ Vehicles are already equipped with black boxes.⁵¹¹ The Nevada and California autonomous vehicle statutes require an autonomous vehicle to store sensor data in read-me format for a time period of thirty seconds prior to a collision.⁵¹² Courts could use this data to determine who was at-fault for the underlying accident, as well as why the crash-optimization algorithm made its decision. If the autonomous technology caused the underlying accident, the manufacturer should be liable, but if the autonomous technology did

⁵⁰⁷ See *Why Ethics Matters*, *supra* note 14, at 72–73 (noting that crash-optimization algorithms could minimize lawsuits and may be a legally and ethically superior strategy).

⁵⁰⁸ See Robert F. Blomquist, *Annoyancetech Vigilante Torts and Policy*, 73 ALB. L. REV. 55, 60 (2009).

⁵⁰⁹ See Michael B. Gallub, *Limiting the Manufacturer's Duty for Subsequent Product Alteration: Three Steps to a Rational Approach*, 16 HOFSTRA L. REV. 361, 365 (1988).

⁵¹⁰ See Goodrich, *supra* note 18, at 289 (“[B]lack boxes serve as a digital transcript of [an] accident.”).

⁵¹¹ *Look, No Hands*, ECONOMIST (Apr. 20, 2013), <http://www.economist.com/news/special-report/21576224-one-day-every-car-may-come-invisible-chauffeur-look-no-hands> (“[C]ars now have a rudimentary version of ‘black box’ data recorders to collect information on the moments just before an accident. Insurers have already begun to offer discounts to motorists who agree to have more sophisticated ones that monitor their driving all the time.”).

⁵¹² CAL. VEH. CODE § 38750(c)(1)(G) (West 2015); NEV. ADMIN. CODE § 482A.190(2)(a) (2015).

not cause the underlying accident, the manufacturer should be dismissed from the lawsuit.

In addition, the government should grant the autonomous vehicle manufacturer immunity from punitive damages, unless the crash optimization algorithm was programmed to make a decision that is completely unreasonable and not merely a difference of ethical opinion. An example of a situation to which a plaintiff should be entitled to punitive damages is in the Shopping Cart Problem if the autonomous vehicle hit and killed the baby because striking the baby carriage would cause less damage to the vehicle than striking the shopping cart. The decision to kill a child to prevent a dent seems completely unreasonable, and society should allow the private tort system to deter such algorithms. Note, however, that if the autonomous vehicle hits the baby carriage because of an error in the crash-optimization algorithm, punitive damages do not seem warranted.

This form of immunity can be applied to the moral dilemmas. In the Trolley Problem, the automaker should receive immunity if it takes control over the vehicle to save the five lives. The bystander is killed because the human operator was negligent, reckless, or intentionally going to kill five persons; that operator caused the need for the algorithm and should be at fault. In other cases, in which a person will not die as a result of the crash-optimization algorithm taking control over the vehicle, society will want to incentivize the manufacturer to do everything it can to save people's lives. Removing liability for the manufacturers would enable them to act. When this happens, the human driver should be civilly responsible for the accident: the driver caused the accident, and her insurance should pay for the harm. In the Tunnel Problem, the autonomous vehicle is not at fault for causing the accident; the child is at fault. Therefore, if the autonomous vehicle hits and kills the child, the automaker should not be civilly responsible. If the automaker programs the vehicle to crash into the side of the wall, the automaker would not be at fault either; the child is at fault. In the Bridge Problem, the manufacturer should not be liable to the operator for driving the vehicle off the bridge, which would, by all means, result in the least amount of total harm. The estate of the operator would be able to sue the bus driver for causing the accident, but it should not be able to sue the manufacturer for programming it to commit self-sacrifice to save so many lives.

VII. CONCLUSION

This article has utilized six moral dilemmas to examine the philosophical and legal questions surrounding crash-optimization algorithms. The programming of crash-optimization algorithms will create difficult ethical questions and ambiguous and troublesome legal responsibility for manufacturers of autonomous vehicles. This article concludes that federal and state governments should provide certainty to autonomous vehicle manufacturers by enacting partial immunity for accidents involving autonomous vehicles.

The programming of autonomous vehicles to act ethically will require collaboration between the government and the car manufacturer. Convincing people to cede control of their lives in a vehicle to a computer will require them to trust the product. For autonomous vehicles to gain acceptance with consumers, the decisions made by a crash-optimization algorithm will require the cooperation of engineers, philosophers, lawyers, and the public. The best way to build that trust is through an open discussion about how the cars will react in ethical situations.⁵¹³ The oncoming discussion of how to implement moral philosophy into autonomous vehicles will provide society with a deeper understanding of each ethical theory and how ethics interacts with the law.⁵¹⁴ And the outcome of these discussions may be that our vehicles make better ethical decisions than we would in similar situations,⁵¹⁵ and this would usher in a new era of road safety in which millions of lives and trillions of dollars would be saved.

⁵¹³ See *Why Ethics Matters*, *supra* note 14, at 82 (“Without looking at ethics, we are driving with one eye closed.”).

⁵¹⁴ See, e.g., Gips, *supra* note 196, at 251–52.

⁵¹⁵ See, e.g., Wallach, *The Challenge of Moral Machines*, *supra* note 3, at 8 (“People are rather imperfect in their ability to act morally. . . . This raises the possibility that artificial entities might be *more* moral than people.”).