2022

Exceptional Driving Principles for Autonomous Vehicles

A. D'Amato

S. Dancel

J. Pilutti

L. Tellis

E. Frascaroli

See next page for additional authors

Follow this and additional works at: https://repository.law.umich.edu/jlm

Part of the Applied Ethics Commons, Science and Technology Law Commons, and the Transportation Law Commons

Recommended Citation

Exceptional Driving Principles for Autonomous Vehicles

Authors

A. D'Amato, S. Dancel, J. Pilutti, L. Tellis, E. Frascaroli, and J. C. Gerdes

EXCEPTIONAL DRIVING PRINCIPLES FOR AUTONOMOUS VEHICLES

A. D'AMATO, S. DANCEL, J. PILUTTI, L. TELLIS, E. FRASCAROLI, and J.C. GERDES

I.   ABSTRACT

Public expectations for automated vehicles span a broad range, from mobility for passengers, to road user safety, to compliance with the traffic code. In most ordinary situations, these expectations can be satisfied simultaneously. But these various expectations can also lead to exceptional scenarios where certain objectives, such as those related to safety, are in tension with road rules. Exceptional driving scenarios challenge motion planning algorithms in automated vehicles to find solutions that are legally grounded, ethically sound, and technically feasible.

The general public's familiarity with exceptional driving scenarios comes from the classic "Trolley Car" problem in philosophy, asking who should live and who should die in an unavoidable collision. These discussions tend to take a consequentialist view by framing the ethical action as the one that achieves the best outcome. By taking a different perspective that views driving as a social contract, the AV's ethical obligations are limited to meeting the duty of care owed to other road users. With this perspective, the existing legal system in the US provides a framework for choosing appropriate behaviors in exceptional driving cases and for answering the Trolley Car problem. This work outlines principles that prioritize care for humans, respect the authority of human-defined traffic law, and ensure that the vehicle avoids decisions that introduce unreasonable risks. Developing AVs that can legally and ethically negotiate exceptional driving scenarios is simply a matter of translating the principles into engineering requirements with no need for new laws or endless philosophical debate.

II.   INTRODUCTION & PROBLEM STATEMENT

In our current transportation ecosystem, operating a motor vehicle and sharing the roadways with other road users entails a certain amount of risk. The legal system helps to manage this risk by placing a duty of care on each road user, including drivers, bicyclists,

pedestrians, etc. and clarifying the expected behavior for tasks such as following, changing lanes, or navigating intersections in specific provisions of the traffic code. Through continual refinement over time, the traffic code and the broader legal system that surrounds it reflect the balance between safety and mobility that society demands.

Developers of automated vehicles must interpret this legal system when designing algorithms that make decisions for any scenario which the vehicle may encounter in its Operational Design Domain. This translation from legal precedent to algorithm is far from trivial. Even straightforward legal structures can be challenging to rigorously code into an algorithm and traffic laws involve a number of subjective concepts related to reasonableness and the extent of the duty of care. For example, the Uniform Vehicle Code's requirement of a "reasonable and prudent" following distance must be translated into a numerical value that the automated vehicle can regulate. Furthermore, such translation must apply to exceptional cases when the duty of care owed to each road user and the provisions of the traffic code cannot be satisfied simultaneously. Developers must handle such cases in a manner that is legally defensible, ethically sound with regards to its treatment of human harm and technically implementable.
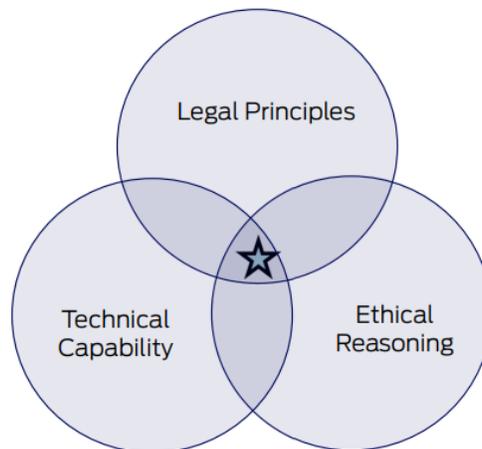
Figure – 1 – The union of these tensions forms the societal expectations for the expected behaviors of automated systems.

Exceptional driving scenarios arise either from a conflict between the duty of care owed to a road user and the provisions of the traffic

code, or from a conflict between the duties owed to multiple road users. This latter set of scenarios includes dilemma situations where collisions are unavoidable regardless of the decision made by the automated vehicle. Such situations can resemble the "trolley car" problem in philosophy, where a person must choose whether to switch a runaway trolley from a track where five people will be killed to another track where one person may be killed.[1] For automated vehicles, an analogous choice may be whether to collide with a pedestrian who steps out into the road immediately ahead of the vehicle or swerve into oncoming traffic. While not a common scenario in practice, such dilemma situations have seen considerable attention in the scientific literature[2] and in popular media and can illustrate differences in ethical perspectives and programming approaches for exceptional driving.

Goodall proposed that dilemma situations be handled by the AV minimizing overall harm, an approach that has often been assumed in popular discussions of trolley car scenarios for AVs.[3] Greene dubbed cars following this approach to be "Utilitarian cars" since they follow the utilitarian philosophical principle that the ethical choice produces an outcome maximizing overall happiness or societal benefit.[4] Although utilitarian cars are straightforward to imagine, in practice they require developers to both accurately determine the likely outcome of a collision scenario and weight the relative merits of these outcomes. The MIT Moral Machine project proposed that society's view of the lesser collision could be crowdsourced by asking participants to choose between pairs of scenarios with different characters losing their lives[5]. Neither the capability to distinguish individuals at this level of granularity nor the ability to accurately predict the outcome of collisions several seconds in the future currently exist.[6] Even if they did, Lin has pointed out that such a strict utilitarian approach could have

---

[1] Judith J. Thomson, *The Trolley Problem*, 94 YALE L. J*.,* 1395, 1395–1415 (1985).

[2] Edward Awad et al*., The Moral Machine Experiment*, 568 NATURE, 59, at 59 (2018); Jean-Francois Bonnefon et al., *The Social Dilemma of Autonomous Vehicles*, 352 SCIENCE 1573, 1573-1576 (2016).

[3] Noah Goodall, *Ethical Decision Making During Automated Vehicle Crashes*, 2424TRANSP. RSCH. REC.: J. TRANSP. RSCH. BOARD*,* 58, at 58 (2014).

[4] Joshua D. Greene, *Our Driverless Dilemma*, 352 SCIENCE 1514, 1514-1515 (2016).

[5] Awad, *supra* note 2.

[6] J. Christian Gerdes & Sarah M. Thornton, *Implementable Ethics for Autonomous Vehicles*, *in* AUTONOMOUS DRIVING 87, 87-102 (J. Christian Gerdes et al., eds., 2013).

unintended consequences such as targeting bicyclists with helmets over those without helmets since they would be more likely to survive a collision.[7] The results also demonstrate how easily societal biases can appear in crowdsourcing (the results from the US demonstrate a preference that fatalities fall on the elderly, the obese and those with criminal records, for instance), calling into question how ethical these outcomes truly are.

As an alternative to crash optimization, other work focuses on defining an appropriate set of rules that can guide vehicle behavior. In sharp contrast to the utilitarian philosophical perspective of optimizing societal outcomes, these approaches take a more deontological approach by bounding the duty or responsibility of the automated vehicle. Responsibility Sensitive Safety (RSS) seeks to define a set of rules that, if universally followed, will result in collision-free driving.[8] The authors propose that the automated vehicle's full responsibility is to follow a collection of rules which include keeping a mathematically defined following distance and lateral distance from other road users and taking an appropriate response when these distances are violated.[9] RSS also requires avoiding collisions with road users who are not following the prescribed proper response as long as this avoidance does not violate the distances to other road users, implementing a form of the duty of care the law requires.[10] A related approach, the Safety Force Field (SFF), also seeks symmetric rules for collision-free driving.[11] The SFF involves each road user claiming space on the road and implementing a driving policy that moves the vehicle away from unsafe sets according to the gradient of a potential function.[12] SFF does not specifically address exceptional driving scenarios when other road users act outside of their range of expected behaviors

---

[7] Patrick Lin, *Why Ethics Matters for Autonomous Cars*, *in* AUTONOMOUS DRIVING 69, 69-85 ((J. Christian Gerdes et al., eds., 2013).

[8] Shai Shalev-Shwartz et al., *On a Formal Model of Safe and Scalable Self-driving Cars*, ARXIV, https://arxiv.org/pdf/1708.06374.pdf (Oct. 27, 2018).

[9] *Id.*

[10] *Id.*

[11] NVIDIA, *The Safety Force Field*, WWW.NVIDIA.COM, https://developer.download.nvidia.com/driveworks/secure/docs/DRIVE_8.0_Relea se_Docs/the-safety-force-field.pdf?g7XRgi0UEfPoK6f65dQoIia9hxirmZuMghroyW8I8X33dnhkKfRueZjqy kCQE96VbpOubZ4-3pmLLB4YM-Zfcx8NW3Uvip8Pgf8zMYnc9zou2yfkzkyAnCaYRVKXHVtkYSWtQA-ofCzXKntTEgUOZ_Y_3dDEama1-XBljPZhYB1HIoNepRTw.

[12] *Id.*

other than emphasizing that the force field and driving policy remain active at all times, regardless of the right of way the traffic code may give the automated vehicle.[13] Neither RSS nor SFF specifically consider exceptional driving scenarios where following the provisions of the traffic code conflict with the goal of avoiding human harm.

Another body of work has examined this possibility of translating the traffic code directly into algorithmic form. Inspired by a complete change to the Dutch traffic code in 1991, den Haan and Brueker proposed a prototype system for coding the law in algorithmic form with the intent of automatically checking the new law for completeness and coherence with previous traffic law.[14] Costescudemonstrated an example of translating Hungarian traffic code requirements for overtaking into requirements for motion planning.[15] While both of these approaches focused on the traffic code itself, Prakken took a broader look at the task of making automated vehicles comply with Dutch traffic law.[16] He observed that conflicts in the rules required explicit prioritization, the semantics of which have been subsequently formalized by Censi et al. in the form of "rulebooks." Less straightforward, however, were questions of liability that required more abstract notions of reasonableness. In these cases, Prakken proposed that industry standards could play a role in adding definition to these concepts for automated vehicles. Finally, Prakken questioned Bonnefont et al.'s assertion that ethical dilemma situations needed to be considered at all for automated vehicles.[17] Since the traffic code and ethics embody the same values of safety and efficiency, he argued, legal and ethical requirements almost completely coincide. Furthermore, he noted, provisions in the law beyond the traffic code itself, such as case law establishing the acceptability of running a red traffic light to get an injured person to the hospital, resolve many apparent conflicts.

---

[13] *Id.*

[14] Den Haan, N., and Breuker, J., "A tractable juridical KBS for applying and teaching traffic regulations." In: Breuker J, Mulder RD, Hage J (eds) *Legal knowledge-based systems. JURIX 91, Model-based reasoning*, Koninklijke Vermande, Lelystad, 5–16, 1991.

[15] Dan M. Costescu, *Keeping the Autonomous Vehicles Accountable: Legal and Logic Analysis on Traffic Code*, *in* Vision Zero for Sustainable Road Safety in Baltic Sea Region 21, 21-33 (Varhelyi et al., eds. 2020).

[16] Henry Prakken, *On the Problem of Making Autonomous Vehicles Conform to Traffic Law*, 25 A.I. and Law, 341, 341–363 (2017).

[17] Bonnefon, *supra* note 2.

This paper begins with the proposition that the traffic code and additional jurisprudence that surrounds it, such as other state laws/regulations and case law, form a social contract for driving. This social contract includes not only the provisions of the traffic code but the larger concept of duty of care and legal principles such as "necessity" required to resolve conflicting objectives. As the product of legislation and judicial interpretation, this social contract embodies the ethical principles deemed important for driving, and therefore, should adequately define both the legal and ethical responsibilities of automated vehicles. By taking a strict interpretation of the duty of care owed to each road user, exceptional driving cases can be handled at a high level by three hierarchical rules, the automated driving system (ADS) shall maintain its duty of care to all road users, the ADS shall actively avoid harm, and the ADS shall follow traffic code This approach not only resolves apparent conflicts between the vehicle code and the desire to reduce harm, but also cleanly resolves dilemma situations. These rules can furthermore be leveraged to develop engineering requirements that take the form of margins necessary to ensure the duty of care. These margins take on a similar form to those of RSS or SFF and can be viewed as minimal values of safety margins necessary to demonstrate reasonable care.

1   LEGAL FRAMEWORK AND EXPECTATIONS

Any set of principles for motion planning in exceptional driving cases must draw from the traffic code relevant to the AV's Operational Design Domain (ODD). Yet traffic codes alone offer insufficient guidance for these cases. At a minimum, concepts in the code intended for human drivers such as "reasonableness" and "due care" require translation into engineering specifications. In some instances, the language of traffic code may lend itself to competing interpretations. Furthermore, human drivers do not always follow traffic code, raising the question of how strictly automated vehicles should follow it.

Given the limitations of traffic code in providing guidance for motion planning, it can be tempting to view the code as only one consideration to balance against other objectives. Doing so, however, ignores the fact that the law provides considerable guidance beyond the traffic code. Traffic and vehicle codes do not exist in a vacuum, but rather, constitute one part of a legal system that includes the judgement of law enforcement, legal principles such

as necessity, and the interpretation and clarification provided by the court system. Taken as a whole, the legal system provides a much more comprehensive framework for choosing the appropriate actions in exceptional driving cases. While developing policies for these cases does still involve some translation of legal principles into engineering specifications, the law offers considerable guidance for this process beyond the traffic code.

1.1   TRAFFIC CODE

In the United States, individual states are responsible for developing vehicle or traffic codes. As a result, each state has its own individual law or code, though all adapt principles—and to a larger degree, wording—from the Uniform Vehicle Code.[18] State traffic codes often appear to be simply a long list of individual rules. However, traffic codes build upon an underlying logic of reciprocal responsibilities designed to enable drivers and other road users to safely share the road.

The Uniform Vehicle Code includes three basic responsibilities – to stay in the lane, allow a reasonable following distance, and drive at a reasonable speed:

1. *A vehicle shall be driven as nearly as practicable entirely within a single lane and shall not be moved from such lane until the driver has first ascertained that such movement can be made with safety.[19]*
2. *The driver of a vehicle shall not follow another vehicle more closely than is reasonable and prudent, having due regard for the speed of such vehicles and the traffic upon and the condition of the highway.[20]*
3. *No person shall drive a vehicle at a speed greater than is reasonable and prudent under the conditions, including actual and potential hazards then existing.[21]*

---

[18] UNIF. VEHICLE CODE (NATIONAL COMM. ON UNIF. TRAFFIC LS. AND ORDINANCE 2000).
[19] *Id.* § 11-309a.
[20] *Id.* § 11-310a.
[21] *Id.* § 11-801.

Approaches to automated vehicle safety such as Mobileye's Responsibility Sensitive Safety,[22] NVIDIA's Safety Force Field,[23] and Motional's Rule Books[24] carry these basic responsibilities from the vehicle code into requirements on the motion planner. The Uniform Vehicle Code augments these three basic responsibilities with other requirements on aspects such as proper turns, overtaking, right-of-way at intersections, and interactions with pedestrians.

In some places, the Uniform Vehicle Code narrows the scope of requirements by establishing a clear hierarchy. For instance, while the code in general prohibits a vehicle from stopping in an intersection or crosswalk, such stops are allowed "*when necessary to avoid conflict with other traffic.*"[25] Implementing this hierarchy in rules in motion planning can be straightforward with techniques such as Rule Books.[26]

In other places, the Uniform Vehicle Code broadens the scope of requirements beyond the specific rules enumerated in the code. The best of example of this broadening is the duty of care drivers owe to pedestrians. The code defines some responsibilities, such as yielding the right-of-way to pedestrians in a crosswalk or sidewalk, explicitly. Yet it clearly broadens the driver's duty beyond this by noting that "*Notwithstanding other provisions of this chapter or the provisions of any local ordinance, every driver of a vehicle shall exercise due care to avoid colliding with any pedestrian or any person propelling a human powered vehicle...*".[27] As discussed later, courts have found that drivers owe a similar duty of care to other road users.

Phrases such as "due care," "reasonable," and "prudent" require some interpretation, and therefore, translating certain written elements of the traffic code into engineering specifications involves some measure of human judgment. These programming decisions should be judged by the same standard of reasonableness applied to human actors. Other prohibitions, such as absolute speed limits in some states, are much clearer. The traffic code itself does not offer

---

[22] Shalev-Shwartz, *supra* note 8.

[23] NVIDIA, *supra* note 11.

[24] Andrea Censi et al., *Liability, Ethics, and Culture-Aware Behavior Specification using Rulebooks*, INT'L. CONF. ON ROBOTICS AND AUTOMATION 8536, 8536-8542 (2019).

[25] U.V.C § 11-1003.

[26] Censi, *supra* note 24.

[27] U.V.C § 11-504.

any justification for driving at a speed above absolute speed limits even if human drivers do so routinely. With the exception of certain situations described in the next section, there is no clear legal defense for violating unambiguous provisions of the traffic code.

The lack of a legal justification for violating specific provisions of the traffic code is not the only compelling reason for designing automated vehicles that strictly comply with the code; compliance also serves a broader public policy objective. As automated vehicles become more common, compliance with traffic code will ensure that changes to that code have a direct impact on vehicle operations and safety. This gives policymakers a more effective mechanism for meeting traffic and safety objectives than setting rules for human drivers who may or may not follow those rules. Designing automated vehicles to comply with traffic codes further ensures a level of human oversight over their operation.

1.2   NECESSITY DEFENSE

Exceptional driving cases can arise in emergency situations when an automated vehicle faces a choice between following the road rules and avoiding a collision. Should an automated vehicle be able to exceed the speed limit to avoid a side impact at an intersection? Should an automated vehicle deviate from its lane to avoid colliding with a pedestrian or bicyclist? Ethical consideration of the "greater good" in these scenarios suggests it would be better to avoid these collisions than to blindly follow traffic rules, creating a tension between the traffic code and the desire to prevent injury or death.

While these situations create tension with the traffic code, they do not, in fact, create tension with the law. Appeals courts have found, for instance, that avoiding human harm can be a defense for speeding or failing to stop and render aid. As Judge Yegan colorfully described in *People v. Morris* , which involved reckless driving on the way to a hospital, "A citizen cannot be reasonably expected to engage in self-sacrifice and bleed to death at the altar of the Vehicle Code by observing the basic speed law and other rules of the road."[28]

To justify such violations of the traffic code in court, the defendant must demonstrate in a legal sense that the violation was necessary to avoid greater harm. The necessity defense is an

---

[28] *People v. Morris*, 191 Cal. App. 3d 8, 8 (Cal. Ct. App. 1987).

affirmative defense, requiring an admission of guilt, but asserting that there was no reasonable alternative to breaking the law. Juries, or a judge as appropriate, must then apply the appropriate criteria for the necessity defense to determine guilt. In some states, such as Texas, necessity is a statutory defense found in the criminal code. In other states, necessity is a common law concept defined through case law. The requirements for the necessity defense in Florida are summarized in pattern jury instructions and provides conditions as follows:[29]

1. The defendant reasonably believed [a danger] [an emergency] existed which was not intentionally caused by [himself] [herself].
2. The [danger] [emergency] threatened significant harm to [himself] [herself] [a third person].
3. The threatened harm must have been real, imminent, and impending.
4. The defendant had no reasonable means to avoid the [danger] [emergency] except by committing the (crime charged) (lesser included offenses).
5. The (crime charged) (lesser included offenses) must have been committed out of necessity to avoid the [danger] [emergency].
6. The harm that the defendant avoided must outweigh the harm caused by committing the (crime charged) (lesser included offenses).

These conditions limit the defense to cases where a significant risk of human harm is imminent, characteristics common to the motivating examples for exceptional driving. They also require that the action is necessary-in the sense that there was no reasonable legal alternative. Finally, the conditions require a balance of harms to demonstrate that the harm avoided was greater than the harm caused. This can be trivially satisfied in cases where the vehicle avoids a collision by causing no harm beyond violating a lane boundary or speed limit. This balance of harms also discourages putting other road users at risk. If the harm caused cannot be reasonably determined (because the automated vehicle changed lanes into an occluded area, for instance), the action can be considered negligent instead of

---

[29] *Criminal Jury Instructions*, THE FLORIDA BAR, https://www.floridabar.org/rules/florida-standard-jury-instructions/criminal-jury-instructions-home/criminal-jury-instructions/sji-criminal-chapter-3/.

necessary. The ultimate determination of whether a driver's actions satisfy the criteria for the necessity defense lies with judge and jury.

When these conditions are satisfied, the law supports violating the traffic code when necessary to avoid collisions, a conclusion that seems both ethical and reasonable. Although other states differ in the exact formulation of the conditions, the principles are similar. Therefore, a legal perspective slightly broader than the traffic code is sufficient to resolve exceptional driving cases that involve a tension between traffic code and human injury.

While the requirement that the defendant did not create the situation is important for the success of the defense, it is not necessarily relevant to an automated vehicle making a decision. Although an automated vehicle later found to have created circumstances in which it violated the traffic code may not be able to claim necessity, that does not imply that the vehicle should favor the traffic code over human injury in these cases. The ethical outcome would still be to avoid human harm even if the vehicle could arguably be found at fault for creating the situation where injury could occur.

1.3   COMMON LAW

As the previous section demonstrated, fully understanding the legal requirements of driving requires looking beyond the traffic code to other sources such as case law and pattern jury instructions which theoretically reflect the law of the state issuing them. These sources provide a more complete sense of the responsibility placed on human drivers. For instance, California's Civil Jury Instructions on the basic standard of care when driving (CACI 700) informs juries that: "A person must use reasonable care in driving a vehicle. Drivers must keep a lookout for pedestrians, obstacles, and other vehicles. They must also control the speed and movement of their vehicles. The failure to use reasonable care in driving a vehicle is negligence."[30] The instructions further point out that this common-law duty extends beyond the vehicle code and cites several relevant cases that further define the control and lookout requirements.[31] In particular, the instructions note that "a driver must at all times exercise ordinary care to avoid a collision including swerving or

---

[30] JUD. COUNCIL OF CAL. CIV. JURY INSTRUCTIONS § 700 (Jud. Council of Cal. 2022).
[31] *Id.*

altering his course, in addition to applying his brakes, if that would be a reasonable means of avoiding the collision."[32]

Appeals court cases can sometimes assist with the task of translating traffic code into engineering requirements. For instance, the Uniform Vehicle Code referenced earlier requires that "*A vehicle shall be driven as nearly as practicable entirely within a single lane*".[33] In turning this requirement into an engineering specification, the first obvious question is how to define "practicable." Texas, which adopts the slightly different phrasing that a driver "shall drive as nearly as practical entirely within a single lane" examined this question in *Leming v. Texas*.[34] In this case, the Court of Criminal Appeals of Texas clarified that: "Failing to stay entirely within a single lane is not an offense if it is prudent to deviate to some degree to avoid colliding with an unexpected fallen branch or a cyclist who has strayed from his bike lane."[35] This case provides considerable clarity to the AV developer since it expressly allows deviating from the lane in two specific situations. Based on this rationale, it seems straightforward to assume that deviating from the lane in similar situations involving pedestrians or other fallen objects would be allowable.

Appeals court cases sometimes clarify that the requirements of the traffic code are not absolute. For example, the rear driver in a rear-end collision in Florida is presumed to be negligent. In *Tozier v. Jarvis*, the judges outline three categories that rebut this presumption – a mechanical failure of the rear vehicle, an unexpected stop or sudden lane change by the front vehicle, and an illegal stop by the front vehicle.[36] In other cases, appeals courts reaffirm that a strict and literal reading of the code is, in fact, correct. Such rulings often appear in challenges to police traffic stops, in which a driver contends that their actions were not a violation, and therefore, the traffic stop (and usually a subsequent discovery of firearms or narcotics) was not legal. Florida courts determined in *State v. Clancey*,) that stopping a vehicle more than 12 inches from a curb was sufficient grounds for a traffic stop, providing a strict interpretation of that requirement.[37]

---

[32] *Guyton v. City of Los Angeles*, 174 Cal.App.2d 354, 362 (1959).

[33] U.V.C. § 11-309a.

[34] *Leming v. Texas*, 493 S.W.3d 552 (Tex. Crim. App. 2016).

[35] *Id.*

[36] *Tozier v. Jarvis*, 469 So. 2d 884 (Fla. Dist. Ct. App. 1985).

[37] *State v. Clancey*, Appeal No. 08-00087APANO (Fla. Cir. App. Ct. August 26, 2009).

There are several challenges when looking for clarification of traffic code in appellate decisions. First, these are state laws and the rulings of a state appellate court holds only for that state. While state appeals courts may refer to cases from other states in making decisions, they are under no obligation to do so. Thus, the conclusions from an appeals court case might not extend beyond roads in that state. Furthermore, very few violations of the traffic code find their way to appellate courts. While most violations go unnoticed, some trigger crashes or citations-a portion of these result in a trial, and a fraction of those decisions are further appealed because of questions about the underlying law. This narrowing at each step of the legal process means that there are many aspects of the traffic code that have not been considered by appellate courts in a particular case. The chances of finding a relevant appeals court case increase in more populous states and with more common incidents (such as rear-end collisions). Where such cases exist, they can clarify motion planning requirements for automated vehicles.

1.4   DRIVING AS A SOCIAL CONTRACT

Several articles such ashave suggested that an automated vehicle has a responsibility to minimize the harmful effects of a collision when a collision becomes unavoidable.[38] This expectation often stems from a discussion of the "Trolley Car" problem in philosophy. In a common statement of this problem, a runaway trolley will crash and kill its five passengers unless a bystander throws a switch to send it down an alternate track, saving the five passengers, but killing another person walking on that track. Philosophers subscribing to a utilitarian view in which the ethical action is whatever maximizes societal good or happiness argue that killing one is better than killing five. Therefore, the bystander should throw the switch. Similarly, faced with an unavoidable collision, they argue that an automated vehicle should seek to minimize harm or damage, and therefore, maximize societal benefits.[39]

Nothing in the traffic code or appellate decisions obligates the automated vehicle (or a human driver, for that matter) to consider the overall societal impact of a crash. Although the necessity defense gives some support for taking actions that can be justified in terms of

---

[38] Shalev-Shwartz, *supra* note 8.
[39] Greene, *supra* note 4.

the greater good, the law simply requires that drivers observe the duty of care owed to other road users and take actions human drivers consider to be reasonable. While a utilitarian approach to philosophy suggests that there is an ethical requirement to consider societal outcomes, the theory of a social contract argues instead that the only ethical requirement is to follow the duties defined by the law.[40] Therefore, designing automated vehicles to reliably observe the duty of care owed to others on the road satisfies both legal requirements and ethical responsibilities from a social contractarian perspective.

The following section contains principles for handling exceptional driving cases based on this understanding of duty of care as a legal and ethical framework. The principles use a strict interpretation that the automated vehicle owes a duty of care to each road user and that it is not allowable to breach the duty owed to one party to achieve better outcomes for another party or society as a whole. This simplifies the balance of harms that needs to be performed for the necessity defense to a determination of whether or not there is a collision. From a philosophical standpoint, this approach is consistent with Judith Jarvis Thomson's conclusions in one of the earliest academic papers to consider the trolley car problem.[41] Citing Ronald Dworkin's observation that "Rights trump utility," she concluded that it was not ethically sound to resolve trolley car problems from a utilitarian perspective if that required violating an individual's rights.[42] Similarly, with the principles presented here, people can rightfully expect that automated vehicles will meet the duty of care owed to each road user and not use them as means to an end. The principles presented here are also consistent with an approach based on virtue ethics which places the virtue of care for others above the virtue of civility, or following the letter of the law.[43]

---

[40] Cudd, Ann and Seena Eftekhari, *Contractarianism*, THE STANFORD ENCYCLOPEDIA OF PHILOSOPHY (2021), https://plato.stanford.edu/archives/win2021/entries/contractarianism/.

[41] Thomson, *supra* note 1.

[42] *Id.*

[43] J. Christian Gerdes, *The Virtues of Automated Vehicle Safety – Mapping Vehicle Safety Approaches to their Underlying Ethical Frameworks*, 2020 IEEE INTELLIGENT VEHICLES SYMPOSIUM, 2020 at 107, 107-113.

2   EXCEPTIONAL DRIVING PRINCIPLES

2.1   PRINCIPLE 0: DUTY OF CARE

> **0. The ADS shall be programmed to maintain a strict duty of care to each road user. The ADS may not violate this duty of care owed to one road user to resolve a conflict with another.**

This initial principle establishes basic expectations for an ADS operated on public roads. While the traffic code in many cases fully defines the duty of care obligations,[44] this principle clearly establishes that developers should not attempt to balance the outcomes of a conflict across actors not involved with the initial conflict.

2.2   PRINCIPLE 1: ACTIVE AVOIDANCE OF HARM

> **1. The ADS shall be programmed to set aside specific provisions of the traffic code when:**
>   a. **it is necessary for abatement and/or avoidance of an imminent collision,**
>   b. **there is no way to avoid the collision while obeying the traffic rules and**
>   c. **the ADS predicts that a greater harm would occur by following the code.**
>   d. **The action does not breach duty of care to another road user**

This principle addresses the cases in which vehicles are confronted with a situation involving potential human harm and the evasive actions that may avoid such harm seem to conflict with certain provisions of the written traffic code. As previously outlined in Section 2, traffic code is only one piece of the larger legal system governing traffic, and the affirmative defense of necessity can provide a framework to resolve this apparent tension between avoiding human harm and written road rules.

This proposal satisfies the essential elements of the necessity defense by recommending that if human harm is imminent and breaking the traffic code is necessary to reduce that harm (where

---

[44] U.V.C, *supra* note 25.

harm is interpreted to mean a collision independent of severity) then the ADS may legally be programmed to break traffic code.

Example:
The ADS may legally be programmed to exceed an absolute speed limit—an action prohibited in traffic code—to avoid an oncoming road user if failure to do so would result in a collision with another road user, and other typically permissible options such as braking would not avoid the collision.
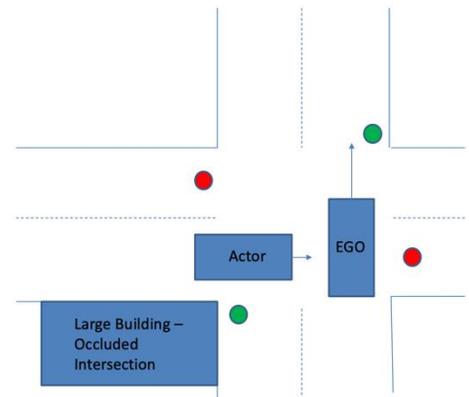

2.3   PRINCIPLE 2A: FOLLOWING EXPLICIT TRAFFIC RULES

2A. Where traffic rules are explicit, the ADS shall be designed to comply with the traffic rules, assuming no conflict with proposal 1.

Unlike situations involving the threat of imminent human harm, there is no clear legal defense for violating unambiguous provisions of the traffic code. This proposal is designed to address the bulk of exceptional scenarios that involve how an ADS operated vehicle should be designed to comply with traffic code. Absent an imminent risk of human harm or interpretable ambiguity within the written code, there is a lack of a legal justification for violating specific provisions of the traffic code; this is the basis on which the proposal is formulated.

Under principle 2A, the ADS shall be designed to execute maneuvers that adhere to written road rules, provided it is not conducting maneuvers necessary for the abatement and/or avoidance of an imminent collision involving harm.

Example:
The ADS shall be designed to travel at or below posted speed limits while executing lane changes, except when avoiding an imminent collision or otherwise as permitted by law.
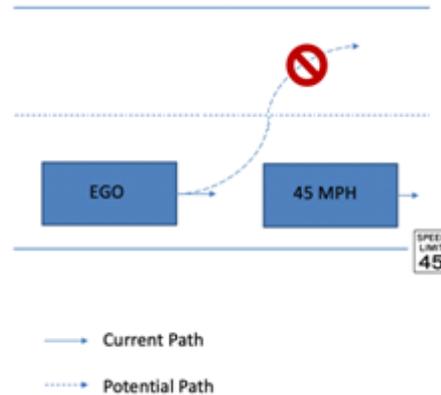
Developers may encounter scenarios where a desired AV behavior is prohibited by law. For example, the desired behavior may be legal in another ODD due to inconsistencies between state road rules. Furthermore, the desired behavior may be exhibited by human drivers. Where the developer believes that a road rule does not improve safety and may inhibit mobility, dialogue with regulators should be established to resolve the conflict.

2.4   PRINCIPLE 2B: INTERPRETING AMBIGUITY IN THE LAW

2B. Where traffic rules require interpretation or judgement, the ADS shall be designed to plan maneuvers that improve its mobility or the mobility of other road users when such maneuvers do not present an unreasonable risk.

While there are many situations in which traffic code is written explicitly, there are also those in which the applicable code is written to incorporate human interpretation or situational judgement. In cases where the law utilizes subjective or interpretable language (such as "practicable" or "when safe to do so"), human road users are expected to use judgement to interpret such language to follow traffic code. In these cases, an ADS may be programmed to conduct such interpretation, or leverage humans to provide that ability.

The ADS shall be programmed using engineering best practices that eliminate unreasonable risk, reasonably interpret written traffic code, interpret certain scenarios that include ambiguous/interpretable traffic code, and/or to leverage human input to interpret such situations and recommend a maneuver.

Example:
In a situation where traffic code states that vehicles are prohibited from crossing into the opposing direction of traffic except when avoiding a *hazard* or *obstruction*, the ADS may be programmed to interpret certain obstacles as *hazards* or *obstructions* (such as a double-parked vehicle) and cross into the opposing lane when reasonably safe to do so to navigate past the obstruction.

Current Path
Potential Path
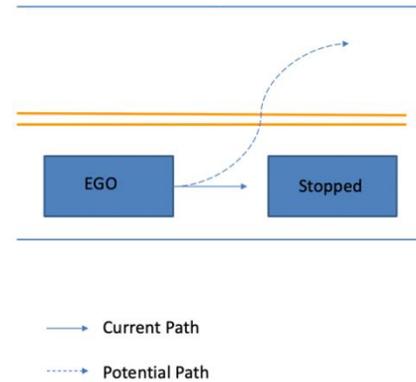
3   IMPLEMENTATION OF EXCEPTIONAL DRIVING PRINCIPLES

Just as these principles for exceptional driving are straightforward to derive from legal principles, they are straightforward to implement in a motion planning algorithm for an automated vehicle. A motion planner uses information about the road ahead, traffic control devices, other road users, obstacles, and occlusions to plan a trajectory for the automated vehicle and determine the acceleration, brake and steering commands necessary to execute this trajectory. The algorithm calculates the desired trajectory over a time horizon which is generally on the order of about 10 seconds. To do this, the planner must make reasonable predictions about the future actions of other road users. Modern motion planners for automated vehicles incorporate some form of optimization to select the best trajectory given a set of desired criteria. These criteria might take the form of hard constraints such as obeying traffic laws or avoiding collision or a cost that should be minimized, such as reducing the acceleration or jerk of the vehicle to improve ride quality for the passengers.[45]

Translating the principles in the previous section into requirements on a motion planner requires defining three concentric regions, or envelopes, around the AV. While the distances that define these envelopes are deterministic, they may not be static. The context

---

[45] S. M. Thornton et al., *Incorporating Ethical Considerations Into Automated Vehicle Control*, 18 IEEE TRANSACTIONS ON INTELL. TRANSP. SYS. 1429, 1429-1439 (2017).

dependencies may include absolute and relative speed of the AV with respect to an object. For example, the envelopes may increase as vehicle speed increases and or in cases where traffic is not flowing at a uniform speed. Other dependencies may include actor type, specifically if the object is a VRU vs another vehicle. Finally, elements including surface conditions, and weather may influence the envelope boundaries. While each of the envelopes may have context dependencies, each of the three envelopes have unique properties that facilitate the execution of the exceptional driving principles, and we describe these properties next.

The ***ride comfort envelope*** represents the region around the AV where the actions of other road users may cause the automated vehicle to execute trajectories that compromise ride quality. Responding to other road users in this region may require braking above a specific threshold or subjectively uncomfortable maneuvers. When objects lie at a distance to the AV beyond this region, the motion planner can accommodate their actions while staying within specified bounds for nominal ride comfort. This region may be designed to consider the ideal attributes or driving style of the brand executing the autonomous driving experience and may be defined implicitly in terms of acceleration limits instead of explicitly in terms of distance.

The next boundary is the ***duty of care envelope***, which is defined as the minimum acceptable contextually dependent distance to another road user/object that satisfies the duty of care obligation. This distance should be informed by the maximum maneuvering capability of the vehicle for the current conditions (e.g., estimated surface friction), the reasonable expectations of other road users and objects, and traffic code. This envelope differs from the ride comfort envelope in that it is an explicitly defined distance bounded by the physics of the vehicle and other actors as opposed to subjective attributes such as ride quality. An actual or predicated violation of this envelope is not necessarily an imminent collision and therefore is insufficient to meet the requirements of Principle 1. When other actors enter the duty of care envelope there is a violation of the duty of care owed by one party to the other.

This boundary is similar to the distances described by Responsibility Sensitive Safety[46] and the Safety Force Field.[47] The safe following distance of RSS, however, is considerably greater than the legal duty of care since it is designed to produce guaranteed collision avoidance. So while the RSS following distance satisfies the duty of care, it also introduces some conservatism. An overly conservative formulation of the duty of care envelope could prevent an AV from taking feasible evasive actions when faced with a situation of conflicting duties to multiple road users.

The third boundary is the ***collision envelope***, which is nominally zero distance from the AV but may be non-zero to include a buffer of uncertainty with respect to the position of an actor or object near the vehicle or sensing limitations. This boundary, like the duty of care envelope, is set objectively. In this instance, sensor types and properties such as field of view are the primary factors in determining these minimum distances. To enable Principle 1, a predicted violation of this envelope is an imminent collision. Figure 2 shows the relationship between the ride comfort, duty of care and collision envelopes.
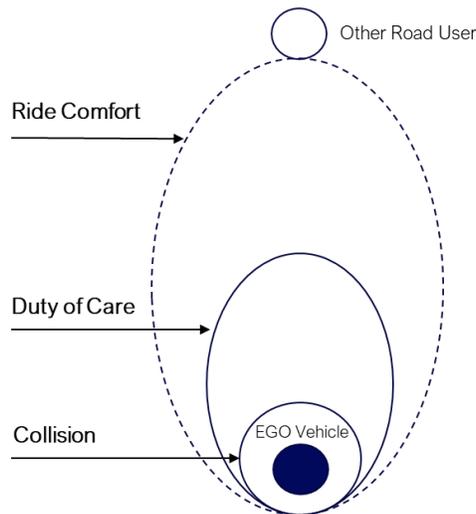
Figure 2 – Example visualization of the various envelopes and considerations for their relative properties.

With the definitions of these three envelopes, the core exceptional driving implementation concept is grounded in the principle that the

---

[46] Shalev-Shwartz, *supra* note 8.
[47] NVIDIA, *supra* note 11.

motion planner shall not, over its planning horizon, execute trajectories that intentionally introduce other actors and/or objects into the duty of care envelope, based on the expectation of reasonable actions of other actors.

Other road users or objects that are involved with either the actual or predicted breach of the duty of care envelope or collision envelope are called the *imminent hazard*. Imminent hazards may arise because of another road user violating *duty of care* to the AV, natural hazards (such as falling objects or animals), or a failure on the part of the AV to adhere to its *duty of care envelope*. In the case where another actor executes an unreasonable action, the AV may not be able to maintain its duty of care envelope to that actor. In these cases, the AV should seek to re-establish its envelope over a reasonable period.
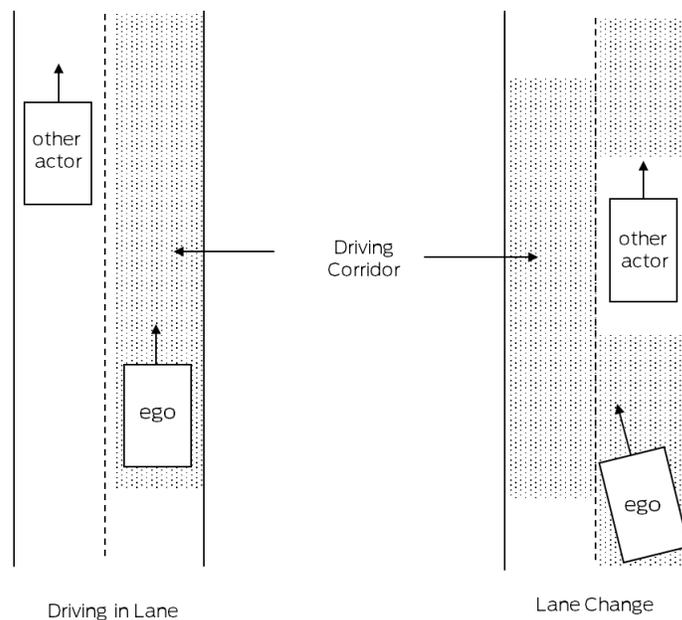
Figure 3 – Example visualization of the driving corridor used by the motion planner to determine violations of the various envelopes as it maneuvers.

Finally, since the ADS may consider multiple trajectories and trajectories are continually updated, the motion planner should assess predicted duty of care violations over the driving corridor, which is the physical space in which the AV intends to travel. The driving corridor corresponds to the physical road lanes when driving within a lane. When changing lanes or turning the driving corridor is the sequence of lanes in which the vehicle intends to travel. Figure 3

shows a visualization of the driving corridor with respect to driving in lane vs a lane change.

With the ride comfort, duty of care and collision envelopes defined along with the notion of an imminent hazard and driving corridor, requirements on the motion planner based on the current operating environment can be organized into four scenarios:

Nominal Conditions

In the nominal use case, where no imminent duty violation is predicted, the motion planner should plan executable trajectories that keep other actors and objects outside the ride comfort envelope, maintain the duty of care envelope, and follow the traffic code. Executable trajectories are the set of positions and velocities of the AV in the driving corridor over a time horizon that accounts for vehicle control and system limitations as known to the motion planner at the time they are determined.

If the motion planner is unable to find an executable trajectory that satisfies these three criteria, then the motion planner shall maintain the duty of care envelope to all road actors and follow traffic code.

Imminent Duty of Care Violation

In the next set of requirements, we introduce an imminent hazard in the driving corridor of the AV, but no collision is anticipated. In this case, the motion planner shall find an executable trajectory that maintains its duty of care to all road users and follows the traffic code. In the event such a trajectory doesn't exist, the motion planner shall choose an executable trajectory that maintains its duty of care to all road users other than the imminent hazard, seeks to reestablish the duty of care envelope to the imminent hazard within a reasonable time and follows the traffic code.

The key element of these requirements is that the AV shall not introduce another road user or object other than the imminent hazard into its duty of care envelope while mitigating the original duty of care violation. A duty of care violation absent prediction of imminent collision requires trajectories that adhere to the traffic code.

Imminent Collisions

The next set of requirements consider the case where there is an actual or predicted violation of the duty of care envelope and it is predicted that the imminent hazard will violate the collision envelope of the AV. In these scenarios, the motion planner should plan executable trajectories that avoid the imminent collision while maintaining the duty of care envelope to all road actors and following the traffic code.

Where such trajectories do not exist, the requirement on maintaining the duty of care envelope to all road actors is relaxed to maintaining the duty of care envelope to all road actors other than the imminent hazard while following traffic code.

If these conditions cannot be satisfied, then Principle 1 allows consideration of executable trajectories that avoid the collision but may violate traffic code. In addition, the motion planner must not introduce actors other than the imminent hazard into its duty of care envelope, and lastly, it must predict that it can either re-establish adherence to traffic code by the end of the trajectory or can achieve a safe state or minimal risk condition.   This avoids generating trajectories where a collision or sustained violation of the law is inevitable, but the motion planner does not recognize this because of the finite time horizon of the planner.
In cases where trajectories do not exist that satisfy these conditions, the motion planner shall plan to use maximum available lateral and longitudinal control to mitigate the imminent collision, while maintaining its duty of care envelope to all road users other than the imminent hazard and concluding the trajectory in a safe state or minimal risk condition.

Faulted State

In the event a failure limits the ability of the motion planner to know its surroundings, the AV must execute an appropriate fallback maneuver. The motion planner should continue to follow the requirements that are currently active preceding the failure as nearly as is practical given the failed state. For example, if at the time of a perception failure an imminent collision was predicted, then that imminent collision and hazard should be assumed to persist even if the location and state of the hazard become unknown.

Example – Pedestrian Interaction

To demonstrate driving outcomes of the exceptional principles, consider the Ego vehicle proceeding straight down a two-lane road divided by a double yellow line (no passing). Suddenly, a pedestrian enters the roadway into the planned driving corridor of the AV. The pedestrian becomes an imminent hazard since the motion planner predicts a duty of care envelope violation will occur and that a collision is imminent. Scenario 1, shown in Figure 4, demonstrates two examples of visualizing driving corridors and the application of the driving principles above reveal the preferred outcome.
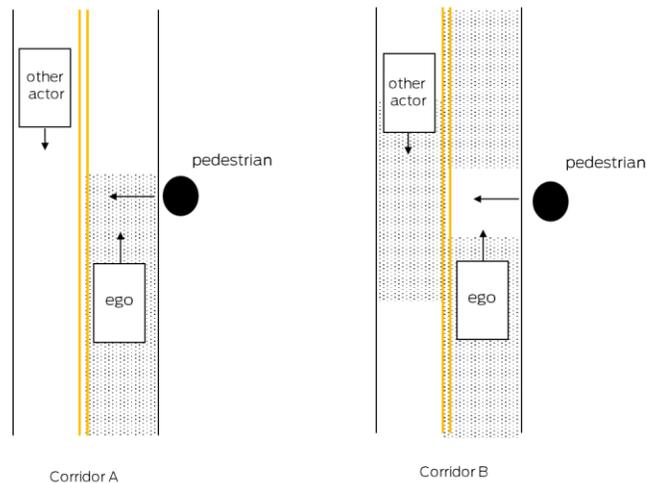

Corridor A   Corridor B
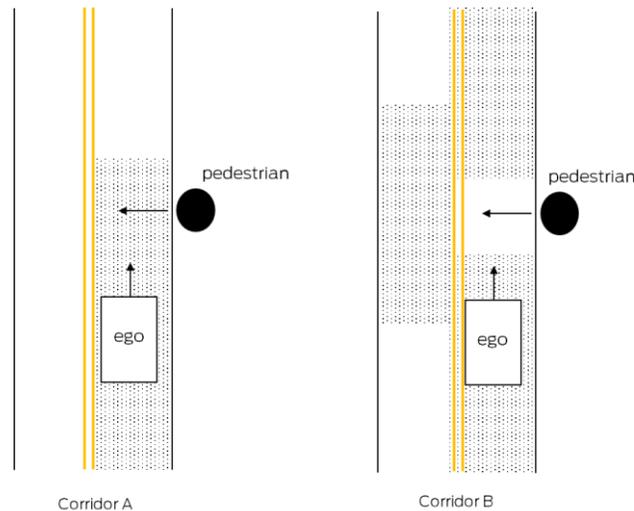
Figure 4 - Scenario 1– example visualization of potential driving corridors in the case where a pedestrian enters the roadway, and no other actors are present. In this scenario, corridor B allows for executable trajectories that avoid the imminent hazard without introducing duty of care violations to other road actors.

Scenario 1 corridor A shows a driving corridor that maintains the lane but results in contact with the pedestrian. Specifically, the ego vehicle would apply maximum braking authority while maintaining the lane, which would result in a collision. In this scenario, there are no other road actors to consider. However, as seen in Scenario 1 corridor B, given the presence of an imminent collision and absence of other road actors, the requirements of Principal 1 are satisfied, and the AV may choose a driving corridor that crosses the double yellow line, breaking traffic code to prevent the collision, but also reestablishing adherence to traffic code at the end of the planning

horizon. Executable trajectories that are within this corridor do not introduce duty of care violations to any other road actor. In this case, corridor B is an appropriate driving corridor.

Next, we consider scenario 2 where another road actor is in the oncoming lane of traffic. Figure 5 is a visualization of Scenario 2 that demonstrates two examples of driving corridors for this scenario. Scenario 2, corridor A shows a driving corridor that maintains the lane. Executable trajectories that are within this corridor use maximum braking authority and steering as needed to maintain the driving corridor and minimize the collision. Corridor A does not introduce duty of care violations to any actor other than the imminent hazard and therefore is the appropriate driving corridor given the example constraints. Scenario 2, corridor B shows a driving corridor that crosses the double yellow line to avoid the collision with the imminent hazard, which in this case, is the pedestrian. However, corridor B does not meet the requirements of Principal 1, since executable trajectories in this driving corridor introduce a duty of care violation to an actor other than the imminent hazard.

Figure 5 – Scenario 2 - example visualization of potential driving corridors in the case where a pedestrian enters the roadway and other actors are present. In this scenario, corridor B violates the duty of care owed to the other actor.

III.   CONCLUSION

Through careful and robust engineering development, the exceptional driving scenarios described in this work are anticipated to be extremely rare events. Despite the low frequency of these events, autonomous vehicle developers should have a strategy in place for prioritization of path planning objectives to ensure consistent vehicle behavior that adheres to established legal frameworks and social contracts. While the principles in this work do not guarantee optimal outcomes, they may engender public trust by ensuring the rights of other road users are respected, and specifically, that AVs and indirectly software engineers are not calculating the worth of one individual vs. another. Furthermore, through these recommendations, AV developers have a tractable solution with respect to the "Trolley Problem".

Some observations on the consequences of these AV principles are:

-The current legal framework and traffic code are complete enough that societal values regarding human harm and compliance with the law can be translated into motion planning requirements.

- Faced with one imminent collision, the AV will not intentionally create another collision to resolve the first. While this does prevent the AV from engaging in crash optimization to improve societal outcomes, it serves to contain the crash by not drawing new actors into the crash scenario.

-The motion planner does not explicitly prioritize one group of road users over another. The Law specifies the duty of care owed to different groups of road users. The motion planner then fulfills the duty owed to each road user.

- Where there is no predicted imminent collision, the AV will follow the rules of the road. As automated vehicles become more common, compliance with traffic code will ensure that changes to that code have a direct impact on vehicle operations and safety. This gives policymakers a more effective mechanism for meeting traffic and safety objectives than setting rules for human drivers who may or may not follow those rules. Designing automated vehicles to comply with traffic codes further ensures a level of human oversight over their operation.

- Strict compliance with traffic laws may compromise mobility. Human drivers do sometimes opt for maneuvers that favor their mobility over compliance with the traffic code. Programming automated vehicles to strictly follow that code may, therefore, not produce the same level of mobility. Where a developer believes that a road rule does not improve safety and may inhibit mobility, a dialogue with regulators should be established to resolve the conflict. This approach provides greater transparency and public input than the alternative of individual AVs or their programmers choosing which laws to follow.

- Motion planners must consider the reasonable actions of other road users in defining their Duty of Care envelope. If the Duty of Care envelope is designed to be overly conservative (by assuming other actors may execute extreme maneuvers, for instance), the AV may lose options to mitigate or avoid a collision that would be available with a smaller envelope. An overly conservative Duty of Care envelope may also compromise mobility. A motion planner that is properly designed according to these principles will never plan a trajectory that breaches the Duty of Care envelope when other actors are behaving within the planner's definition of reasonable.

- Different Safety Envelopes such as RSS and SFF have been proposed in the literature. A Safety Envelope may have different properties than a Duty of Care Envelope so long as the Safety Envelope is greater than or equal to the Duty of Care Envelope under all circumstances. A larger Safety Envelope raises the same concerns as the conservative Duty of Care Envelope discussed above.