UC DAVIS LAW REVIEW ONLINE

VOL. 51 MAY 2018

Three Types of Structural Discrimination Introduced by Autonomous Vehicles

Hin-Yan Liu[*]

The advent of autonomous vehicles has been hailed by commentators as introducing an improvement for traffic safety by promising to reduce the overall number of road accidents as the technology matures. Some advocates even hint at a moral imperative to structure incentives and smooth over barriers in order to induce widespread usage of autonomous vehicle in order to actualize this potential. The orientation towards safety concerns, however, foregrounds the debate on crash-optimization and imports the trolley-problem thought-experiments to the question of autonomous vehicles.

 Associate Professor, Centre for International Law, Conflict and Crisis, and Coordinator, Artificial Intelligence and Legal Disruption Research Group, Faculty of Law, University of Copenhagen. Email: hin-yan.liu@jur.ku.dk. I would like to thank Senior Online Editor Lugar Sungil Choi for his insightful comments and engaging discussions that greatly improved this Article. This is an updated and broadened version of an argument first aired at RoboPhilosophy 2016 at the University of Aarhus, Hin-Yan Liu, *Structural Discrimination and Autonomous Vehicles: Immunity Devices, Trump Cards and Crash Optimisation*, *in* WHAT SOCIAL ROBOTS CAN AND SHOULD DO 164 (Johanna Seibt, Marco Nørskov, & Søren Schack Andersen eds., 2016), and subsequently elaborated in Hin-Yan Liu, *Irresponsibilities, Inequalities and Injustice for Autonomous Vehicles*, 19 ETHICS & INFO. TECH. 193 (2017).

This paper examines the potential for three types of structural discrimination to be woven into the fabric of these developments. First, given the emphasis placed upon decisional agency by trolley-problem scenarios, there will be systematic privileging of the occupant vis-à-vis pedestrians and other third-parties. Second, there is the prospect for structural discrimination arising from the coordinated modes of autonomous vehicle behavior that is prescribed by its code, or which is converged upon through learning algorithms operating towards similar goals and within similar constraints. This leads to the potential for hitherto individuated outcomes to be networked and thereby multiplied across fleets of vehicles. The aggregated effects of such algorithmic policy preferences will thus cumulate in the reallocation of benefits and burdens to certain categories of persons in a relatively stable manner. This in turn raises the spectre of a more pernicious form of active structural discrimination where the possibility of crash-optimization casts a protective shield over certain individuals at the cost of third-parties. Third, the introduction of autonomous vehicles within the framework of crash-optimization will likely precipitate infrastructural changes, which in a literal sense, will introduce or exacerbate structural forms of discrimination with regard to human access to public space.

TABLE OF CONTENTS

INTRODUCTION

Autonomous vehicles are now reaching the statistical mileage that human drivers on average cause a fatality,[1] and have reportedly killed a pedestrian on public roads.[2] While human statistics may not be indicative of the standard for autonomous vehicle performance, this unfortunate milestone raises an opportunity to discuss the potential for improved road safety that an integrated system incorporating such vehicles might allow. Some commentators imply a societal benefit of introducing autonomous vehicles,[3] while others suggest that there might be moral reasons to tailor regulation that is conducive to the development of such vehicles if there are good reasons to favor their introduction.[4]

While there are significant technical factors and statistical evidence suggesting that autonomous vehicles will be capable of superior performance in driving activities compared to human drivers, the rush towards discussions of crash-optimization should be tempered with considerations for unintended consequences. Patrick Lin has provided an excellent analyses of many relevant ethical issues associated with crash-optimization and the application of trolley-problem ethics to the functioning of autonomous vehicles.[5] The idea of crash-optimization is

---

[1] *See* Martin Robbins, *Statistically, Self-Driving Cars Are About to Kill Someone. What Happens Next?*, GUARDIAN (June 14, 2016, 5:39 AM), https://www.theguardian.com/science/2016/jun/14/statistically-self-driving-cars-are-about-to-kill-someone-what-happens-next (noting that it takes one hundred million miles for a human driver to cause a fatality in the United States).

[2] *See* Ian Bogost, *Can You Sue a Robocar?*, ATLANTIC (Mar. 28, 2018), https://www.theatlantic.com/technology/archive/2018/03/can-you-sue-a-robocar/556007/ (reporting the death of a pedestrian by Uber's autonomous vehicle in Tempe, Florida).

[3] *See* Gary E. Marchant & Rachel A. Lindor, *The Coming Collision Between Autonomous Vehicles and the Liability System*, 52 SANTA CLARA L. REV. 1321, 1322 (2012) (implying a widespread use of autonomous vehicles if they reduce the frequency and severity of accidents); *see also Federal Automated Vehicles Policy: Accelerating the Next Revolution in Roadway Safety*, NAT'L HIGHWAY TRAFFIC SAFETY ADMIN. (2016), https://one.nhtsa.gov/nhtsa/av/pdf/Federal_Automated_Vehicles_Policy.pdf.

[4] *See* Alexander Hevelke & Julian Nida-Rümelin, *Responsibility for Crashes of Autonomous Vehicles: An Ethical Analysis*, 21 SCI. & ENGINEERING ETHICS 619, 621 (2015) ("If there are good *moral* reasons for finding the . . . introduction of autonomous cars to be desirable, this can produce a moral obligation for the state to fashion the legal responsibility for crashes of autonomous cars in a way which helps the development . . . of autonomous cars.").

[5] Patrick Lin, *Why Ethics Matters for Autonomous Cars*, *in* AUTONOMES FAHREN 69 (Markus Maurer et al. eds., 2015) [hereinafter *Why Ethics Matters*]; *see also* Patrick Lin, *Here's a Terrible Idea: Robot Cars with Adjustable Ethics Settings*, WIRED (Aug. 18, 2014, 6:30 AM), http://www.wired.com/2014/08/heres-a-terrible-idea-robot-cars-with-adjustable-ethics-settings/ [hereinafter *Adjustable Ethics Settings*]; Patrick Lin, *The Robot Car of*

essentially that the overall damage that ensues from an unavoidable crash, determined from an objective standpoint, could be limited and even consciously minimized.[6] This framing imports clear analogies with the much-discussed trolley-problem thought-experiments.[7] At root, the trolley-problem presents variations of a restrained choice: undertake an action that will result in quantitatively less harm, or passively allow events to unfold that will result in greater objective harm.[8] Recent applications of the trolley-problem to the introduction of autonomous vehicles have sparked a debate about how to appropriately program such vehicles to function in precisely such exigencies and by extension how to minimize unavoidable harms more generally.[9] Given the projections of autonomous vehicle performance, the implementation of layered redundancies, and the prospect for constant networked communication in a broader traffic system dominated by autonomous vehicles, engineers project that unavoidable crash scenarios will be rare and decreasing phenomena.[10]

Tomorrow May Just be Programmed to Hit You, WIRED (May 6, 2014, 2:42 PM), http://www.wired.com/2014/05/the-robot-car-of-tomorrow-might-just-be-programmed-to-hit-you/ [hereinafter *The Robot Car*].

[6] *See* Lin, *Why Ethics Matters, supra* note 5, at 72 ("Some accidents are unavoidable — such as when an animal or pedestrian darts out in front of your moving car — and therefore autonomous cars will need to engage in crash-*optimization* as well. Optimizing crashes means to choose the course of action that will likely lead to the least amount of harm, and this could mean a forced choice between two evils . . . ." (emphasis in original)); *see also* Lin, *Adjustable Ethics Settings*, *supra* note 5 (explaining the trolley-problem as a moral dilemma, in that it is generally better to harm fewer people than more, to have one person die instead of five).

[7] *See generally* DAVID EDMONDS, WOULD YOU KILL THE FAT MAN? THE TROLLEY PROBLEM AND WHAT YOUR ANSWER TELLS US ABOUT RIGHT AND WRONG (2013) (discussing the trolley-problem through the history of moral philosophy).

[8] For an overview, see *id.*

[9] *See, e.g.*, Lin, *Why Ethics Matters*, *supra* note 5; Jean-François Bonnefon et al., *The Social Dilemma of Autonomous Vehicles*, SCI., June 2016, at 1573, http://science.sciencemag.org/content/352/6293/1573.full; Lauren Cassani Davis, *Would You Pull the Trolley Switch? Does it Matter?*, ATLANTIC (Oct. 9, 2015) http://www.theatlantic.com/technology/archive/2015/10/trolley-problem-history-psychology-morality-driverless-cars/409732/; Cory Doctorow, *The Problem with Self-Driving Cars: Who Controls the Code?*, GUARDIAN (Dec. 23, 2015, 7:00 AM) http://www.theguardian.com/technology/2015/dec/23/the-problem-with-self-driving-cars-who-controls-the-code; Patrick Lin, *The Ethics of Autonomous Cars*, ATLANTIC (Oct. 8, 2013), http://www.theatlantic.com/technology/archive/2013/10/the-ethics-of-autonomous-cars/280360/.

[10] *See* Aarian Marshall, *To Save the Most Lives, Deploy (Imperfect) Self-Driving Cars ASAP*, WIRED (Nov. 7, 2017, 12:01 AM), https://www.wired.com/story/self-driving-cars-rand-report/. *But see* Peter Hancock, *Are Autonomous Cars Really Safer Than Human Drivers?*, CONVERSATION (Feb. 2, 2018, 6:29 AM),

These practical assertions need not detract from the considerations elaborated upon in this paper, however, because the prospect for accidents, by definition, can never be entirely eliminated. Furthermore, the debate on crash-optimization and the application of trolley-problem ethics to autonomous vehicle operations constitutes an attempt to engage with the moral, social, and legal implications that such technologies introduce, as well as highlight the alignment of interests undergirding the widespread commercial introduction of a new mode of transport.

It is interesting that Patrick Lin transforms the impulse towards crash-optimization into targeting practices, leading to discussions and burgeoning empirical studies as to whether autonomous vehicles should be programed to kill at least in constrained contexts.[11] The question remains, however, whether results of rule-based and sequentially cumulative decisions, which the current autonomous vehicles must comply with, can amount to targeting and thus harming individuals per se. There are three reasons for this. First, targeting implies a direct and intentional action against a pre-identifiable target, while trolley-problem style accident scenarios involve coerced, time-restricted choices that curtail the scope of volition and furthermore are not necessarily aimed at harming a particular victim. Second, targeting is to a large extent decontextualized and preordained, while trolley-problem style accident scenarios set strong situational contexts which constrain decisions by placing parameters upon the range of possible actions. Third, unlike targeting, autonomous vehicle programming does not determine particular future outcomes, but instead establishes probabilistic courses of action in given contexts. The point here is that the programming of the autonomous vehicle only makes certain outcomes more likely than others, but does not determine which outcome manifests. Taken together, there is significant conceptual distance between the intentional, unconstrained, and directly causal act of targeting and the restrained time-pressured dilemma under which a decision in a trolley-problem style accident scenario takes place.

Yet, while crash-optimization need not necessarily lead to targeting for these reasons, it is important to note that such practices are by no means excluded, and such stark prospects are explored in more detail in the hypothetical outlined below. Instead, attempts towards crash-optimization may precipitate unintended consequences, which being

http://theconversation.com/are-autonomous-cars-really-safer-than-human-drivers-90202.

[11] *See* Bonnefon et al., *supra* note 9.

likely to become aligned with commercial and personal interests, would be difficult to foreclose or dispel.

I. PRIORITIZING THE OCCUPANT IN AUTONOMOUS VEHICLES THROUGH TROLLEY-PROBLEM SCENARIOS

Before moving to discuss the structural biases embodied in crash-optimization impulses, an under-explored dimension of applying trolley-problem ethics needs to be explored which is the foundation of the first form of structural discrimination introduced by autonomous vehicles.[12] This concerns the contextualization of autonomous vehicles within trolley-problem ethics, and the legitimating function played by that thought-experiment in validating crash-optimization impulses that privilege the perspective and interests of the occupant over competing concerns, such as third-party bystanders.

The ethical vantage point for the original trolley-problems placed the decision-maker outside of the scenario entirely (pulling levers to divert the trolley, or to release the trap-door under a fat-man who will fall onto the tracks and stop the trolley in its tracks).[13] Thus, in the original trolley-problem scenarios, the decision-maker was disinterested because she was insulated from both the benefits and the consequences of her decision: this abstracted position arguably undergirds the ethical nature of the conundrums.

This independence and impartiality of the decision-maker was subsequently lost when trolley-problem ethics were applied to autonomous vehicle scenarios because the decision-maker, be it the manufacturer[14] or the occupant,[15] now has a vested stake in the outcome. Furthermore, this stake in the outcome is both practical and immediate — for the manufacturer, profits are aligned with serving the exclusive interests of the customer[16] and for the occupant,

[12] I owe the elaboration of this section to Lugar Sungil Choi.

[13] *See* EDMONDS, *supra* note 7, at 140.

[14] *See, e.g.*, David Z. Morris, *Mercedes-Benz's Self-Driving Cars Would Choose Passenger Lives Over Bystanders*, FORTUNE (Oct. 15, 2016), http://fortune.com/2016/10/15/mercedes-self-driving-car-ethics/.

[15] *See* Lin, *Why Ethics Matters*, *supra* note 5; *see also* Giuseppe Contissa et al., *The Ethical Knob: Ethically-Customisable Automated Vehicles and the Law*, 25 ARTIFICIAL INTELL. & L. 365 (2017) (arguing for equipping autonomous vehicles with a device enabling the occupants to ethically customize their vehicles to choose between different settings corresponding to different moral approaches in unavoidable accident scenarios).

[16] Recall that corporations are legally-mandated to maximize profits and their shareholder values and have been implicated in externalizing costs on a systematic basis. *See* JOEL BAKAN, THE CORPORATION: THE PATHOLOGICAL PURSUIT OF PROFIT AND

personal safety is an obvious concern despite the possibility to inculcate one's ethical preferences into an autonomous vehicle.[17] Transferring the ethical vantage-point of the decision-maker in trolley-problem scenarios to the manufacturers or occupants of autonomous vehicles risks placing the dilemma into the hands of those who have active interests in the outcome. This distorts the ethical underpinning of the original thought-experiment, rendering it a legitimation strategy for the defense of narrow personal and corporate interests at the expense of decisions that would maximize public benefit.

In present context, framing crash-optimization as trolley-problems for autonomous vehicles privileges their occupants over pedestrians and other third-parties.[18] The distortion of trolley-problem ethics through decisions of the beneficiaries might therefore transform the very nature of the trolley-problem from an objectively-based crash-optimization calculus to one that is subjective and centered upon the occupant's perspective. Thus, the question is transformed from which pedestrian is the least worth saving (or the most worth saving) to *which pedestrian would minimize harm to the passenger*.[19] This is a very different question to that posed by original trolley-problem scenarios. Yet such a conclusion is justified by the logic of those ethical thought-experiments.

The systematic privileging of the occupant is further exacerbated by the sublimation of responsibility for the outcomes that are precipitated by autonomous vehicle behavior.[20] The lack of responsibility practices, capable of consistently identifying, confronting, and addressing such problems, will further obscure this type of structured discrimination. Further opacity is introduced by the difficulty of establishing the ground upon which discrimination is disallowed — while many human rights instruments leave the prohibited grounds of discrimination open-ended and non-exhaustive,[21] access to

POWER 102 (2005).

[17] *See* Contissa et al., *supra* note 15.

[18] This is because the corporate interests of the manufacturers converge with those of the occupants, whom, whether as purchasers of autonomous vehicle services or products, are the driving force behind the profitability of those goods and services.

[19] Patrick Lin touches upon how prioritizing the passengers would lead to certain types of logics in the pedestrians it targets. For example, protecting occupants would justify the strategy of colliding with the lightest objects and thus systematically burdening children with higher levels of risk. *See* Lin, *Why Ethics Matters*, *supra* note 5, at 72.

[20] *See* Hin-Yan Liu, *Irresponsibilities, Inequalities and Injustice for Autonomous Vehicles*, 19 ETHICS & INFO. TECH. 193, 194-200 (2017).

[21] Such grounds are provided in Article 26 of the International Covenant on Civil

autonomous vehicles is hardly an innate or immutable characteristic that should warrant such protection from a formally legal perspective. This converges with the difficulties encountered with asserting responsibilities for autonomous vehicles to subtly shift burdens of risk to pedestrians and to other third-parties.

Public policy must be developed to counter the collusion of narrow private interests that configure structures of privilege enjoyed by those with access to autonomous vehicles, especially given the difficulties of identifying and challenging this reallocation of the risks and costs of autonomous vehicles to those who are not directly benefitting from them. A starting point may be to democratize the trolley-problem thought-experiment as it is applied to autonomous vehicles,[22] and to aggregate ethical positions taken from the perspectives not only of the agent-occupier, but also of the patient-pedestrian and other third-parties. Rather than let the manufacturers and occupants impose their ethical (or practical) preferences unhindered,[23] it would be wise for society as a whole to debate: first, the impetus towards crash-optimization and second, the ensuing questions of how to implement and monitor the configuration to be adopted. But since crash-optimization and trolley-problem scenarios, as applied to autonomous vehicles, take the thought-experiments beyond abstract and disinterested deliberations, an aggregated approach incorporating the interests of pedestrians and other third-parties would also be necessary. This is the rough equivalent of soliciting the opinions of the workers who are tied to the tracks in the original trolley-problems as to what the ethically correct course of action should be.

II. STRUCTURAL BIASES IN CRASH-OPTIMIZATION AND TROLLEY-PROBLEM ETHICS

The second type of structural discrimination introduced by the autonomous vehicle concerns patterned outcomes as consequences of the structured accumulation of hitherto reactive, individuated, and

and Political Rights ("ICCPR"), which provides that: "All persons are equal before the law and are entitled without any discrimination to the equal protection of the law. In this respect, the law shall prohibit any discrimination and guarantee to all persons equal and effective protection against discrimination on any ground such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status." Art. 26, G.A. Res. 2200 (XXI) A, International Covenant on Civil and Political Rights (Dec. 16, 1966).

[22] For an example akin to this point, see MORAL MACHINE, http://moralmachine.mit.edu/ (last visited Apr. 16, 2018).

[23] *See* Contissa et al., *supra* note 15; Morris, *supra* note 14.

isolated decisions in crash-optimization contexts. The terminological framing of *autonomous* vehicles, prejudices towards the continuation of discreteness as the analytical lens. In other words, discussing *autonomy* in vehicles smooths analogies towards contemporary anthropocentric driving practices where un-networked human beings operate vehicles according to individualized training and decision-making processes. While such an extension of contemporary practices appears obvious and unassailable, it is by no means necessary nor inevitable. Instead, the patterned outcomes will arise through communication and coordination (in a network or system) or convergence (through learning algorithms tasked towards achieving certain similar goals within similar environmental constraints).

Thus, a potentially dystopian narrative can be woven by the trinity of: the safety improvements promised by the introduction of the autonomous vehicle;[24] the crash-optimization debate as strained through trolley-problem ethics;[25] and the cumulative effects of centralized, coordinated, or converging tendencies of algorithmic policy preferences. Assessing the projected safety benefits of introducing autonomous vehicles will be an empirical exercise hinged to their performance and thus beyond the scope of this paper. Thus, the second form of structural discrimination results from aggregated and accumulated outcomes built upon discrete and independent events presented in trolley-problem scenarios. Taken together, these independent events strongly indicate a disposition towards aggregated outcomes that are structural biased and which verge upon systematic discrimination.

Turning first to the distorting effects of straining the behavior of autonomous vehicles through the prism of trolley-problem ethics, the important characteristics are that the consideration is adopted from the perspective of the isolated individual actor.[26] The frame, therefore, focuses the ethical conundrum upon what course of action should be adopted in a given situation, diverts attention away from the impact of

[24] *See* Marshall, *supra* note 10 ("[T]ens or even hundreds of thousands of lives could be saved by self-driving cars, even if regulators allow less-than-perfect cars on the road.").

[25] *See* Lin, *Why Ethics Matters*, *supra* note 5.

[26] The actor's perspective is important because trolley-problem ethics are built upon the choices made by those possessing agency in relation to the autonomous vehicle: in other words placing the ethicist in the "driver's seat." Trolley-problem ethics would look rather different if the decisions were made from the perspective of those tied to the tracks, those deciding upon public policy for how society should solve trolley-problems, or those who are assigned to mitigate or insure against such risks.

those decisions upon third-parties, and obscures the prospect of cumulative effects. In other words, the trolley-problem's myopic focus upon the causes and consequences of a single scenario overlooks the possibility that the results will yield patterns of externalities and induce trends that cannot be predicted by studying the ethical justifiability of the decision in isolation.[27] A rough analogy can be drawn with the concept of emergence to illustrate these effects — complex and unforeseen consequences can arise through accretion from mere adherence to simple rules.[28] Studying those rules or their application in isolation would not yield indications of complexity, let alone be able to account for the resultant intricacies. Similarly, studying trolley-problem ethics will not indicate the broader impacts and effects of accumulating a succession of individually justifiable decisions. It may be that individually justifiable decisions may, in aggregation, result in disastrous outcomes.

The concern raised about aggregated outcomes becomes important because of the potential for autonomous vehicles to operate upon similar, or even identical, algorithms.[29] In other words, the coding for autonomous vehicles could be compiled in a centralized manner by the developer or manufacturer and then broadly distributed: the effect being to multiply any algorithmic policy preferences that might have been implicitly or explicitly incorporated. Even the simplest of rules written into the code for autonomous vehicles could cascade a range of relatively consistent but unforeseen outcomes depending on the situation.[30] Patterns emerge when such outcomes are aggregated, and these patterns are indicative of structural biases in the ensuing allocation of burdens and benefits in relation to the functioning of the autonomous vehicles. Because this variant of structural discrimination arises primarily from the hand-coded algorithms, it may be easier to rectify since the structural discriminatory effects arise primarily from magnifying and spreading its effects.

---

[27] Again, that the consideration of the decision-maker's perspective is problematic because it assumes the autonomous vehicle ethics should be descended from those who drive. Even when ethical considerations are removed, however, there are pragmatic commercial incentives towards saving the occupants over the pedestrians for vehicle manufacturers who are financially rewarded for favoring the lives of the occupant-purchasers.

[28] *See* JAMES GLEICK, CHAOS: MAKING A NEW SCIENCE 9-32 (1997); MELANIE MITCHELL, COMPLEXITY: A GUIDED TOUR 56-70 (2011).

[29] This claim is grounded on the likelihood that manufacturers will converge upon certain types of autonomous vehicle platforms, analogous to the Microsoft/Macintosh platforms.

[30] GLEICK, *supra* note 28, at 9-32; MITCHELL, *supra* note 28, at 56-70.

Alternatively, absent centralization, learning algorithms may converge upon types of behavior that are adaptively optimized to reach certain goals in the context of given constraints. Much of the present hype around artificial intelligence involves the contemporary advances in machine learning, and in particular its subset of deep learning, which have proven to be immensely successful in solving persistent problems such as image recognition.[31] Machine learning is where an algorithm is no longer hand coded such that its output is pre-determined.[32] Rather machine learning algorithms possess flexibility in their capacity to adapt and improve with iterative feedback.[33] In Tom Mitchell's broad definition of learning: "A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P*, if its performance at task *T*, as measured by *P*, improves with experience *E*."[34] In this sense, machine learning is thus really an optimization process which improves iteratively without necessarily being taught what to do.

Similar effects can then occur through cascades in communication, where members of a fleet or brand of autonomous vehicles then share the winning strategies arrived at by one of its members. But with the case of learning algorithms, this convergence upon a consistent and repetitive pattern of behavior need not even be based upon communication and coordination (that is to say that communication and coordination can give rise to structural discrimination, but are not necessary in contexts involving learning algorithms). This is because the optimization function of learning algorithms will converge upon patterns of behavior where the specified goal and the environmental constraints are similar (and will likely do so in unforeseeable ways).[35]

This feeds into another blind spot of trolley-problem ethics. The calculus is conducted with seemingly featureless and identical "human units," as the variable being emphasized is the quantity of harm rather than its character or nature.[36] The remedy to this broad equality is

[31] *See generally* IAN GOODFELLOW ET AL., DEEP LEARNING (2016) (explaining the history, the overview, and the modern application of "deep learning").

[32] *See generally id.*

[33] *See generally id.*

[34] TOM MITCHELL, MACHINE LEARNING 2 (1997).

[35] An intuitive way to grasp this is to consider the visual aesthetics of Google's Deep Dream Generator. While the generated works appear to be unique, an identifiable aesthetic emerges from repetitive overall patterns that make each work conform to a certain style. *See* DEEP DREAM GENERATOR, https://deepdreamgenerator.com (last visited Apr. 16, 2018).

[36] The set-up for the original trolley-problem dilemmas make it quite clear that the relevant dimensions are of action and inaction, and the number of human beings who are in harm's way. *See generally* EDMONDS, *supra* note 7.

found in the non-identity problem which underscores the fact that different lives are saved and sacrificed. This fact is well-hidden in assertions that there will be a net saving of lives usually expressed in statistics.[37] The non-identity problem infuses individuality into otherwise utilitarian considerations and thus brings trolley-problem considerations closer to real-world scenarios.

The prospect of recognizing individual characteristics, however, threatens to open the floodgates concerning which features and whose interests can and will be recognized and what weighting these should be accorded. To some extent this has already taken place where gestures were made towards a dilemma between diverting an autonomous vehicle to hit either a motorcyclist who wears a helmet or another who does not. As Patrick Lin observes, adherence to a principle aimed at minimizing harm would result in the autonomous vehicle hitting the helmeted motorcyclist because her odds of surviving the accident are higher, yet such an outcome places burdens upon exactly those who adopted prudential measures to minimize their exposure to risk.[38] This example neatly encapsulates the structural concerns of centralized preferences and risk allocation because it illustrates the potential for certain groups to consistently bear a greater burden by factoring their characteristics into the utilitarian crash-optimization framework.

But the non-identity problem, by eroding the staunch equality of individuals as expressed in the trolley-problem, imports the prospect of privileging certain characteristics or penalizing others in the pursuit of crash-optimization. At the level of the individual decision, the effects of which are isolated from each other, resolving the dilemma in either direction may remain ethically justifiable and socially acceptable because the ethically correct course of action remains sufficiently contentious.

The issue that the non-identity problem raises in the context of autonomous vehicles is not so much that individual lives may either be saved or lost as a direct consequence of introducing autonomous vehicles, but rather that the same centralized and coordinated rules govern those decisions. As a result, the prospect for the same algorithmic preferences controlling the vehicles to be replicated across any number of such vehicles leads to the possibility for identical responses that are governed by the same rule-structure. This in turn

[37] *See* Patrick Lin, *The Ethics of Saving Lives With Autonomous Cars Is Far Murkier Than You Think*, WIRED (July 30, 2013, 6:30 AM), http://www.wired.com/2013/07/the-surprising-ethics-of-robot-cars/.

[38] *See* Lin, *Why Ethics Matters*, *supra* note 5, at 73-74.

creates a systemic and collective dimension whereby the generated outcomes will be reliably and systematically skewed according to the coded preferences, whether intentional or not. The crucial differentiator is thus the removal of hitherto discrete and independent actions undertaken by individuals and the range and diversity of available responses that flow as a result.[39] The subsequent harmonization in accumulating these responses skews together results in systematic biases in relation to certain sets of characteristics. If the preferred or penalized preferences map onto individual or group characteristics for which discriminating based on those characteristics is impermissible,[40] these structured biases have been translated into systematic discrimination.

Thus systematic discrimination *looks very much like discrimination*, but cannot readily be challenged as breaching legal anti-discrimination provisions as such. This is because the outcomes are patterned through a consistent aggregation of small and discrete decisions that are biased in a certain direction, but may not reflect discriminatory intent. In other words, the aggregated outcomes converge upon the outcome-space that maps onto discrimination, but legal recognition for such effects as discrimination are unlikely to be forthcoming. A variant of this argument has been elaborated upon by Scott Veitch, when he asks rhetorically: "What registers in law as a wrong? This question has a deceptively simple answer: what registers in law as a wrong is a breach of the law."[41]

Difficulties lie ahead for those seeking to challenge this potential for systematically discriminatory outcomes. The framework of the trolley-problem is misaligned with the realities of autonomous vehicle operation by foregrounding isolated individual decisions made from the agent-occupier perspective, as discussed above, while overlooking more subtle cumulative impact of those actions. As the technology is

[39] The idea that individuals would meaningfully "decide" upon a course of action during the split-seconds that precede an accident might be idealizing such scenarios. The point, however, under crash-optimization logics is two-fold: first that the course of action is determined beforehand such that it is proactive rather than reactive; and second that the course of actions is patterned and cumulative rather than discrete and unconnected.

[40] *See* Art. 26, G.A. Res. 2200 (XXI) A, International Covenant on Civil and Political Rights (Dec. 16, 1966).

[41] SCOTT VEITCH, LAW AND IRRESPONSIBILITY: ON THE LEGITIMATION OF HUMAN SUFFERING 92 (2007). Veitch continues: "Strictly speaking, then, it is not any particular loss — physical suffering, economic harm, or whatever — that registers as such, but rather only that suffering or harm is given legal cognisance. If there is suffering that involves no breach of the law, there is no wrong." *Id.*

capable of placing identical programming in a range of vehicles, or where learning algorithms arrive at similar strategies given the goals sought and constraints imposed, the impact of those decisions become readily aggregated. Yet, algorithmic preferences can be defended on the basis that each individual decision is justifiable so long as the overall policies are designed to minimize overall harm since any discriminatory effect is seemingly tangential due to a lack of intention (and possibly also foreseeability).[42]

Thus, cumulative impacts remain legally unrecognized. Objections to claims of structural discrimination can be readily resisted because of the difficulties associated with articulating the harm in theory and demonstrating the harm in practice. To illustrate the liminal position of such structural discrimination, consider the fundamental human right to equal protection enshrined in Article 26 of the International Covenant on Civil and Political Rights: "the law shall *prohibit* any discrimination and *guarantee to all persons equal and effective protection* against discrimination [on non-exhaustive grounds]."[43] Structural discrimination as identified and elaborated here appears both to fall outside the scope of this provision while being simultaneously incongruous to the protections that it affords.

III. INTENTIONAL DISCRIMINATION AND THE IMMUNITY DEVICE THOUGHT-EXPERIMENT

While any discriminatory consequences emerging from these processes will be structural and consistent in nature, at this point they remain passive, implicit, and even unintended. The concern is that the

[42] In the context of structural discrimination in the context of urban design, the need to demonstrate discriminatory intent has proved to be a difficult hurdle to clear. *See* Sarah Schindler, *Architectural Exclusion: Discrimination and Segregation Through Physical Design of the Built Environment*, 124 YALE L. J. 1836, 1979-87 (2015) (holding that "[t]he plaintiff must show that he himself is injured by [defendant's discriminatory act]" to trigger strict scrutiny (quoting Vill. of Arlington Heights v. Metro. Hous. Dev. Corp., 429 U.S. 252, 265 (1977)). This position was subsequently confirmed in *Memphis v. Greene*, where even clear disparate impact was insufficient to ground a claim of discriminatory intent. 451 U.S. 100, 128-29 (1981). The lack of discriminatory intent may scupper any and all attempts to ground claims of discrimination where AI is concerned for the simple observation that AI do not "intend" anything. If clear disparate impact is insufficient to overcome the discriminatory intent requirement, it is foreseeable that decisions made and influenced by AIs would be largely insulated from legal allegations and challenges underpinned by discrimination.

[43] *See* Art. 26, G.A. Res. 2200 (XXI) A, International Covenant on Civil and Political Rights (Dec. 16, 1966) (emphasis added).

crash-optimization discourse, when coupled with the non-identity problem, paves the way towards seemingly justifiable, even permissible, forms of active and intentional discrimination. Making this leap involves flipping the utilitarian calculus — rather than attempting to minimize harm in unavoidable situations, that aim instead will be to maximize the collective well-being of society.

What if the decision as to whom the autonomous vehicle were to hit included considerations of an individual's positive traits, such as innate talent, cultured ability, or latent potential? As Joseph Louis Lagrange quipped after Antoine-Laurent de Lavoisier, widely acknowledged as the father of modern chemistry, was guillotined: "It took them only an instant to cut off this head, and one hundred years might not suffice to reproduce its like."[44] In this vein, would society not have an interest, albeit a potentially perverse one, in programming its autonomous vehicles to preserve the lives of its prized scientific and cultural elite in pursuit of the public benefit and greater good? Establishing protective preferences for certain categories of persons is implicit in the emotional appeal when school buses are inserted into the trolley-problem dilemma. The presence of children invokes our intuitive responses that the lives of the young and innocent are intrinsically worth protecting, yet the impermissibility of discrimination on the basis of age has been highlighted in this context.[45] This emphasizes the distance between the ethical and legal quandaries that arise where policies are anchored in actions that would be permissible, or at least justifiable, as individual actions. In Patrick Lin's example,[46] an individual choosing to allocate the risk to the elderly woman in a trolley-problem scenario could justify their actions on a range of socially acceptable reasons. Elevating such an individual preference to a collective system that repeatedly allocates risk along such lines, however, would constitute systematic discrimination on the grounds of sex and age that would be legally impermissible and broadly undesirable. Thus, individualized ethics may legitimize defensible yet short-sighted policies that make for terrible cumulative consequences.

Extending the question of protective preferences beyond the tenuously justifiable grounds of merit and individual capacity, what about policies that protect individuals with social or political status, or financial capital? At the moment, our political leaders and diplomatic

[44] RALPH HERMON MAJOR, THE DOCTOR EXPLAINS 134-35 (1931).

[45] *See* Lin, *Why Ethics Matters*, *supra* note 5, at 70-71.

[46] *See id.*

representatives clearly enjoy high levels of personal protection, which is not broadly provided to the ordinary citizen, and those with the financial resources are permitted to purchase additional security. In the former situation, it is arguable that the extra protection merely offsets the additional risks borne by those leading public lives and those who hold offices serving the public good. At this stage at least, it appears that any policy preferences which are implemented in favor of exceptional and important individuals, whom society has demonstrated vested interests in protecting, is consistent with the current policies. In the latter situation, however, the arrangement may simply be one of markets and commodities, with security being purchased as a tradable good like any other. Here the lines of justifiability begin to blur because policy preferences become unhinged to the broader public benefit. Moreover, trolley-problem crash scenarios invoke a zero-sum game — the preference to protect one party in the dilemma diverts the risk to the other party. While the contemporary purchasing of private security is not unproblematic, it remains arguable that an individual purchasing additional security does not directly displace dangers to other parties as a result of this decision. In other words, direct externalities are generated by decisions in trolley-problem scenarios that need to be factored into the holistic calculus when considering which individuals require the protection for societal well-being.

Pursuing this logic to accord greater protection to certain individuals or groups in society, it would not be implausible or unreasonable for a large entity, like the manufacturers of autonomous vehicles, to issue what I would call here an "immunity device" — the bearer of such a device would become immune to collisions with autonomous vehicles. With the ubiquity of smart personal communication devices, it would not be difficult to develop a transmitting device to this end which signals the identity of its owner. Such an amulet would protect its owner in situations where an autonomous vehicle finds itself careening towards her. It would have the effect of deflecting the car away from that individual and thereby divert the car to engage in a new trolley-problem style dilemma elsewhere. If the justifications for the bearer of the immunity device are sufficiently strong, and their numbers suitably restricted, this might be a practical response to the new quandaries introduced by autonomous vehicles. After all, a human being behind the wheel in an identical situation would likely make the same decision, for example, if presented with the choice and ability to hit an ordinary member of

the public to avoid killing a Nobel laureate or an architectural genius[47] in an unavoidable crash.

But this appears to be the thinnest end of a very large wedge. The scenario introduced above is binary and absolute because the immunity device offers complete protection. Yet, if the technology becomes available it is difficult to see how proliferation can be contained. Developing the immunity device would introduce the ability for cars to communicate with their potential victims' devices in the event of an "accident" (now a metaphorical term since eventualities are calculated), and a range of pressures would push these capabilities into desperate or greedy hands. For obvious functional reasons, however, a widespread system of immunity devices would be impracticable and self-defeating. Instead, hierarchical nuances would have to be introduced — essentially a status ranking system that eases the trolley-problem calculus. Individuals occupying public spaces would essentially be playing a game of trump cards with each other in the event of a trolley-problem involving an autonomous vehicle. Whoever bears the highest status aversion device would be spared at the expense of those who possess lower status ones. This would rapidly become an uncomfortable outcome for what initially appeared to be a satisfactory configuration of benefits and burdens imposed by the autonomous vehicle of the future.

At this point, we can reinsert the targeting issue that Patrick Lin suggested as the corollary of crash-optimization. It might be that cascading trump card calculations will become both overly complex and unnecessarily convoluted to achieve the desired outcome. As a heuristic to the endeavor, the solution may simply be for an autonomous vehicle to target the lowest value individual in range of the unavoidable collision. It would appear that the consequence of conducting a cascade of trolley-problem dilemmas through a series of trump card comparisons would ultimately lead to the result of the individual bearing the lowest value being hit anyway: so why not short circuit the whole process? Where the system of risk allocation is hinged to the individual's value to society, or at least perceived to reflect this ranking, such a targeting policy appears to equate to causing the least amount of objectively determined damage to society and thereby becomes justifiable.

This then asks the question as to how to allocate the particular status an individual should possess, and therefore the concomitant

[47] Appropriate in the trolley-problem context because Antoni Gaudi is perhaps the most famous fatality involving an actual urban tram. *See* EDMONDS, *supra* note 7, at 53.

level of risk that she should bear in relation to autonomous vehicles. All its variants run against the bold proclamation in the Universal Declaration of Human Rights that "[a]ll human beings are born free and equal in dignity and rights."[48] Were a meritocratic system to be implemented, we might end up with the more palatable situation sketched above. More likely, however, given the commercial context of developing and marketing autonomous vehicles, is that the distribution will be made on economical grounds given the financial opportunities for profit-making that such a system would introduce. It is easy to envisage existing customer loyalty or benefit programs that extend into this realm when offered as additional security or safety. Similarly, insurance companies might market such a system of devices in order to cover the potential decline of revenue in insuring human drivers that would be precipitated by the widespread introduction of autonomous vehicles.

It remains possible, however, that if the immunity devices were to be distributed by a public authority, this would enable that authority to nudge behavior accordingly.[49] While such nudging can be used to incentivize laudable behaviors, for example rewarding those with low carbon footprints or those who have remained outside of the penal system, there is also the more dystopian prospect for the immunity devices to be allocated according to an individual's social credit.[50] In other words, the ability to centralize the allocation of risk and reward in the context of autonomous vehicles enables this scheme to be rolled into broader incentive structures and nudging practices exercised by public authorities. The involvement of public authorities, however, could formally push these issues into the realm of administrative law and human rights, which might hold the potential for greater scrutiny. That said, the difficulty with demonstrating discriminatory intent,

[48] Art. 1, G.A. Res. 217 A, Universal Declaration of Human Rights (Dec. 10, 1948).

[49] My thanks to Lugar Sungil Choi again for making me elaborate upon this point. On nudging, see generally CASS R. SUNSTEIN, THE ETHICS OF INFLUENCE: GOVERNMENT IN THE AGE OF BEHAVIORAL SCIENCE (2016); RICHARD H. THALER & CASS R. SUNSTEIN, NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS (2008); Cass R. Sunstein, *The Ethics of Nudging*, 32 YALE J. ON REG. 413 (2015).

[50] *See, e.g.*, Rachel Botsman, *Big Data Meets Big Brother as China Moves to Rate Its Citizens*, WIRED (Oct. 21, 2017), http://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion ("[I]n China . . . the government is developing the Social Credit System (SCS) to rate the trustworthiness of 1.3 billion citizens."); Mara Hvistendahl, *Inside China's Vast New Experiment in Social Ranking*, WIRED (Dec. 14, 2017, 6:00 AM), https://www.wired.com/story/age-of-social-credit/ ("[I]n 2014, the Chinese government announced it was developing what it called a system of 'social credit.'").

especially in the context of protecting virtuous individuals (the zero-sum displacement of risk upon others being largely overlooked) would most likely render these additional tools useless.

Turning to the market for trading autonomous vehicle trump card ratings, a particularly interesting scenario would be where the trading took place in a closed market. In such a scenario, it is the relative differential that will be valued and effective — individuals seeking a protective advantage would need to possess a higher rating in comparison to others within a closed system. A zero-sum game would thus be implemented where individuals seeking high status would be required to purchase the differential protection from other members within the same society, thereby instigating a competitive and polarizing scheme. This shows the potential for rapid evolution from a protective system initiated to safeguard societal interests into a pure market system that thrives on exaggerating the differences between members of the same society.

Beyond marketing such devices, their allocation could run along the lines of rewards and punishments, akin to the social credit system discussed above. In addition to receiving public honors, for example, one's relative value in the protective schema might be enhanced to reflect society's increased interest in the individual. Conversely, it might even be possible to allocate additional risks as a form of punishment for societal transgressions — those who have acted to the detriment of society might be forced to bear a burden of increased risk as a form of restitution or compensation. These may be more justifiable, and societally aligned, ways of developing such a scheme. Yet, however the final structure may be, its hierarchical effects are evident, as are the tendencies towards heuristics, categorization, and entrenchment. And again, given the commercial context, these societally aligned schemes might, at best, operate in tandem with the market-based approach in a blended hybrid system.

Leaving aside the practical details associated with the future market for autonomous vehicles — such as whether manufacturers will offer such systems, whether they would be preferential towards their own clientele, and whether they would coordinate and centralize such a scheme — the ramification of such a system is that risk of injury and death shifts towards prevention rather than cure. Rather than enforcing accident insurance upon human drivers to cover for their future faults, this system would instead place a large share of the burden upon the ordinary citizen to avoid or minimize the likelihood of injury arising from autonomous vehicles. This relocates burdens away from the beneficiaries of autonomous vehicles and their use to

bystanders who do not necessarily have a stake in the introduction of these technologies.[51] In pre-determining the allocation of risks and costs in advance of any accidents, the remaining fault that must be covered by the occupier (and in reality be passed on to the manufacturer of the autonomous vehicle) would be for departures from the course of action that has been promised in advance, and not for the actual outcome caused by the autonomous vehicle.

IV. STRUCTURAL DISCRIMINATION IN THE CORPORATE PROFIT-DRIVEN CONTEXT

To make matters even more complex, autonomous vehicles are not being developed from any semblance of a transparent and neutral situation, but are rather driven forward by private corporate entities seeking to make material profits from their efforts and investments.[52] Even if such motives do not taint the product directly, preferences that increase profits are likely to become embedded in the decisional architecture of the autonomous vehicle that a company produces.[53] Profitability and client interests collide in such instances to externalize the risks to third-parties.[54] It is neither uncommon nor unreasonable for a vehicle manufacturer today to emphasize the safety features of its

[51] In a different context, but converging upon similar conclusions, are the objections that the introduction of autonomous vehicles would propagate existing discrimination and increase social segregation, and that it would marginalize human beings from time-honored public infrastructure. *See* Ian Bogost, *How Driverless Cars Will Change the Feel of Cities*, ATLANTIC (Nov. 15, 2017), https://www.theatlantic.com/technology/archive/2017/11/life-in-a-driverless-city/545822/; Ian Bogost, *Will Robocars Kick Humans Off City Streets?*, ATLANTIC (June 23, 2016), https://www.theatlantic.com/technology/archive/2016/06/robocars-only/488129/ [hereinafter *Will Robocars*].

[52] *See* Bogost, *Will Robocars*, *supra* note 51 ("No matter the scenario, one thing's for sure: When Silicon Valley runs our automobiles, they'll do so according to the business practices of the technology industry, not according to the principles of local urban planning and civic life.").

[53] *See generally* BAKAN, *supra* note 16.

[54] For example, the background to the recent Uber self-driving car fatality lies in an Executive Order made by Arizona Governor Doug Ducey to undertake "any necessary steps to support the testing and operation of self-driving cars on public roads within Arizona." Ariz. Exec. Order 2015–09 (Aug. 25, 2018), https://azgovernor.gov/file/2660/download?token=nLkPLRi1; Bogost, *supra* note 2. Later, on March 1, 2018, Ducey updated that order to allow fully autonomous driving provided that a "minimal risk condition" (in turn meaning a "reasonably safe state . . . upon experiencing a failure") was met by the autonomous vehicle's systems. Under this direction, Arizona residents have been subjected to heighten risks associated with traffic as a result of being a real-world test-bed for the application of autonomous vehicles. Ariz. Exec. Order 2018–04 (Mar. 1, 2018), https://azgovernor.gov/file/12514/download?token=6jUxyR_C; Bogost, *supra* note 2.

models that protect its customers and passengers.[55] And a human driver who takes the decision to hurt others in order to protect her own interests similarly acts within the boundaries of social acceptability.[56] These tendencies might imperceptibly be extended into the realm of autonomous vehicles without much commentary, deliberation, or evaluation. Yet, when these two characteristics are united in the autonomous vehicle, the realities may shift. The vehicle subtly transitions between being an artifact towards being an agent — from something that is inanimate and subject to human control to something that observes, orients, decides, and acts. Crossing this line implicates programming that may systematically elevate the safety of its customers and occupants over all others, not least because an autonomous vehicle would not be able to convincingly make account of its decision-making processes. In an important sense, the outcome, where the prospect of harm is unavoidable, has not only been automated and pre-determined, but these results also become multiplied and systematic.

The private developmental context also imposes restrictions in terms of accessibility and influence over prescriptive policies. Not only do the technical dimension of the underlying technology create barriers that curtail widespread engagement with determining the eventual behavior of autonomous vehicles, but the usual political tools deployed for oversight and accountability are incapable of penetrating the corporation. A prescient warning can be found in *Wisconsin v. Loomis*,[57] where a six-year prison sentence was based in part upon the report generated by a secret algorithm: because COMPAS was a private company's proprietary software,[58] the petition was based upon the inability of the defendant to inspect or challenge this process.[59] The

[55] *See, e.g.*, Marshall, *supra* note 10; Morris, *supra* note 14.

[56] A point of distinction might subsist in the fact that the absence of the very possibility of responsibility for autonomous vehicles means that decisions for how these should behave need to be specified up front. This is unlike the traditional driving scenario whereby human drivers are held to account in retrospect for their decisions and actions, a possibility which may justify the discretional latitude which human drivers enjoy that might remain out-of-bounds for autonomous vehicles.

[57] 881 N.W.2d 749 (Wis. Sup. Ct. 2016).

[58] *See id.* at 761 ("[T]he developer of COMPAS, considers COMPAS a proprietary instrument and a trade secret. Accordingly, it does not disclose how the risk scores are determined or how the factors are weighed."); *see also COMPAS Classification*, EQUIVANT, http://www.equivant.com/solutions/inmate-classification (last visited Apr. 17, 2018).

[59] *See Loomis*, 881 N.W.2d at 757 ("[Defendant] asserts that the . . . use of a COMPAS risk assessment at sentencing violates a defendant's right to due process . . . because the proprietary nature of COMPAS prevents him from assessing its accuracy."); *see also id.* at

opaque nature of the corporation also presents significant obstacles for those seeking to challenge how autonomous vehicles will operate, and limits the opportunities for increasingly marginalized third-parties to raise their concerns.

To make a legal point here, these technical and formal barriers may lead to injustice insofar as harms and damages remain unrecognized as legal wrongs or injuries. The notion of wrong implies an infraction with potential moral, and possibly legal, consequences and begins to assert the need for accountability. Injury as understood in its root sense of *injuria*, meaning an invasion of another's rights or conversely a legally actionable wrong, demonstrates the distance between the harm and its (legal) recognition.[60] Harms and damages, on the other hand, need not bear ethical opprobrium nor legal repercussion unless these are recognized and transformed into their ethical and legal categories.[61] This gap obscures the prospect of damage occurring without injury being recognized,[62] and this may occur as a direct consequence of third-parties being systematically disadvantaged by the algorithmic preferences being unable to review or challenge those veiled policies. Taken together, both the opportunities to influence the behavior of autonomous vehicles and the ability to hold the conduct to account become severely curtailed.

In situations where the occupant of the autonomous vehicle (whether termed as a pilot, operator, occupant, passenger, or something else entirely) is imputed with responsibility for autonomous vehicle accidents, the specter of scapegoating emerges. This is both because the human beings in such positions do not possess the control, predictability, and foresight required to reliably and effectively alter the course of autonomous vehicle behavior; and because human beings can only be held in a role-responsibility sense of fulfilling obligations, which have been decoupled to the causal-

761 ("[Defendant] argues that he is in the best position to refute or explain the COMPAS risk assessment, but cannot do so based solely on a review of the scores as reflected in the bar charts . . . [U]nless he can review how the factors are weighed and how risk scores are determined, the accuracy of the COMPAS assessment cannot be verified." (footnote omitted)). For commentary on the *Loomis* case, see Adam Liptak, *Sent to Prison by a Software Program's Secret Algorithms*, N.Y. TIMES (May 1, 2017), https://www.nytimes.com/2017/05/01/us/politics/sent-to-prison-by-a-software-programs-secret-algorithms.html; Frank Pasquale, *Secret Algorithms Threaten the Rule of Law*, MIT TECH. REV. (June 1, 2017), https://www.technologyreview.com/s/608011/secret-algorithms-threaten-the-rule-of-law/.

60 *See* VEITCH, *supra* note 41, at 85-92.

61 *See id.*

62 *See id.*

responsibility that the unfortunate consequences demand.[63] Where the autonomous vehicle is equipped with warnings and other means of reverting control back to human drivers, where technology failures or accidents appear imminent,[64] this might merely insert human beings as moral crumple zones[65] in human-robot interaction systems.[66] As long as responsibility for accidents remains legally ambiguous, it will be practically expedient and materially profitable to displace liability upon human beings who are in the liminal zone between simultaneously driving and not driving the (autonomous) vehicle.[67] But this is a different way the law fails to recognize the injury within the damage, discussed above,[68] that would perpetuate the very sources for the lack of this legal recognition. The use of proximate human beings as moral crumple zones that absorb legal liability for autonomous vehicles may be a less pressing, but still important, form

[63] Along the lines of the development of the two different types of responsibility issues, see Hin-Yan Liu, *Refining Responsibility: Differentiating Two Types of Responsibility Issues Raised by Autonomous Weapons Systems*, *in* AUTONOMOUS WEAPONS SYSTEMS: LAW, ETHICS, POLICY 325, 329 (Nehal Bhuta et al. eds., 2016) ("The responsibility of the . . . commander is both limited by [her] lack of control and foresight (circumstantial issues) and characterized by the need to fulfil the obligations that attach to . . . [her] role (conceptual issues). . . . [T]he commander . . . can either argue that the outcomes were not foreseeable or . . . assert that [the commander] had discharged [her] role responsibilities.").

[64] *See, e.g.*, Brian Fung, *The Driver Who Died in a Tesla Crash Using Autopilot Ignored at Least 7 Safety Warnings*, WASH. POST (June 20, 2017), https://www.washingtonpost.com/news/the-switch/wp/2017/06/20/the-driver-who-died-in-a-tesla-crash-using-autopilot-ignored-7-safety-warnings/.

[65] *See* M.C. Elish, *Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction* 3 (We Robot 2016, Data & Soc'y Research Inst., Working Paper, 2016), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2757236 (defining "moral crumple zone" to describe the result of ambiguity within systems of distributed control due to the incongruities between control and responsibility. Incongruities exist because control has become distributed across multiple actors while the social and legal conceptions of responsibility remained generally about an individual).

[66] *See id.* at 19-22 (explaining that in the instance when accidents are caused by a miscommunication between a driverless car and a human-driven car, "responsibility is shifted to other drivers on the road, and these drivers become the moral crumple zone, taking responsibility for a failure where, in fact, control over the situation is shared").

[67] For example, "the letter of the Arizona executive order seems to suggest that the human operator is on the hook for any traffic infractions while he or she is in the vehicle, even if it's in fully autonomous mode." Bogost, *supra* note 2; *see also* Ariz. Exec. Order 2018–04 (Mar. 1, 2018) ("[T]he person . . . operating the fully autonomous vehicle may be issued a traffic citation or other applicable penalty in the event the vehicle fails to comply with traffic and/or motor vehicle laws.").

[68] *See supra* notes 63–64 and accompanying text.

of structural discrimination, since its presence will consistently shield the corporations that design and deploy the technology from public scrutiny and legal liability, as well as the more distal human beings who benefit financially from the operation of autonomous vehicles.

V. STRUCTURAL DISCRIMINATION IN URBAN DESIGN AND LAW REVISITED

The lack of clear and unambiguous liability structures dilutes incentives towards producing autonomous vehicles which are safe not only for their occupants but also for the public at large. The discriminatory potentials embodied in the autonomous vehicle and the systems which will grow around in support of their operation, however, has broadly evaded the radar. Thus, the incentives towards creating non-discriminatory autonomous vehicle systems are immature. Given the prospect for public-private partnerships to form around the deployment of autonomous vehicle systems as surrogates for truly public transportation systems,[69] a more traditional form of structural discrimination may arise. This is due to the segregating effects of such technology that privilege wealth and education, and the predicted transformation of urban areas through the process embedding autonomous vehicles and their supporting systems.[70]

Thus, beyond the aggregated effects of crash-optimization calculations is a form of structural discrimination that may arise and be woven into the very fabric of urban space.[71] Sarah Schindler has contextualized the concept of "architectural exclusion," which involves "the exclusionary effects of . . . seemingly innocuous features of the built environment."[72] It might be asserted, if Lawrence Lessig is

[69] *See* Bogost, *supra* note 2 ("Given sufficient economic incentive to pursue public-private partnerships between municipalities and technology companies . . . states might choose to adopt industry-friendly regulatory policy in exchange for changes to the urban environment."); *see also* Cecilia Kang, *Pittsburgh Welcomed Uber's Driverless Car Experiment. Not Anymore.*, N.Y. TIMES (May 21, 2017), https://www.nytimes.com/2017/05/21/technology/pittsburgh-ubers-driverless-car-experiment.html.

[70] *See* Bogost, *Will Robocars*, *supra* note 51 ("Once non-local technology companies have a direct impact on the financing and planning of urban spaces, the negative impact on less white, less wealthy communities could be severe.").

[71] Schindler, *supra* note 42, at 1943 ("[M]onumental structures of concrete and steel embody a systematic social inequality . . . that, after a time, becomes just another part of the landscape." (quoting Langdon Winner, *Do Artifacts Have Politics?*, 109 DAEDALUS 121, 124 (1980))).

[72] *Id.* at 1939-40.

correct that architecture constitutes a modality of regulation,[73] that *structural* discrimination can literally be built into the fabric of urban space.[74] Unfortunately, because law here is relegated to a regulatory modality on par with architectural regulation, their parity as regulatory modalities suggests that legal remedies will be ineffectual against structural discrimination as embodied in urban space. This is perhaps why the most that can be said of such practices is that they are forms of architectural *exclusion*, as opposed to "architectural discrimination," which would legalistically be non sequitur.

Indeed, the facilitation and impediment with regard to transportation are well-documented strategies of architectural exclusion, and ready analogies can be drawn with precedents found in public transit and road infrastructure.[75] The introduction of autonomous vehicles will add nothing new in this sense, but merely exacerbate existing trends of architectural exclusion. Perhaps the most famous example is of Robert Moses and the bridges over the parkways of Long Island which are "extraordinarily low." In the words of Langdon Winner:

> It turns out, however, that the two hundred or so low-hanging overpasses on Long Island were deliberately designed to achieve a particular social effect . . . . Automobile-owning whites of "upper" and "comfortable middle" classes, as he called them, would be free to use the parkways for recreation and commuting. Poor people and blacks, who normally use public transit, were kept off the roads because the twelve-foot tall buses could not get through the overpasses. One consequence was to limit access of racial minorities and low-income groups to Jones Beach, Moses's widely acclaimed *public* park.[76]

---

[73] *See id.* at 1940 (explaining Lawrence Lessig's regulatory theory that behavior may be regulated or constrained, in part, by architecture); *see also* Lawrence Lessig, *The New Chicago School*, 27 J. LEG. STUD. 661, 662-63 (1998) (arguing that, as one of the four modalities of regulation, architecture can "restrict and enable in a way that directs or affects behavior"); Lawrence Lessig, *The Law of the Horse: What Cyber Law Might Teach*, 113 HARV. L. REV. 501, 507 (1999) (arguing that architecture is a "fourth feature of real space that regulates behavior").

[74] *See* Schindler, *supra* note 42 at 1953-73 (providing examples of architectural exclusion).

[75] *See id.* at 1960-72 (showing that architectural exclusion comes in the form of decisions about where to place transit stops; highway routes; bridge exits; road infrastructure; one-way, dead-end, and curvy streets; and road signs).

[76] Langdon Winner, *Do Artifacts Have Politics?*, 109 DAEDALUS 121, 124-25 (1980)

What is particularly astounding is the permanence of architectural regulation, and in this sense, exclusion — "[f]or generations after Moses has gone and the alliances he forged have fallen apart, his public works, especially the highways and bridges he built to favor the use of the automobile over the development of mass transit, will continue to shape [New York City]."[77] The longevity of architectural exclusion might suggest its place along constitutional provisions as literally *entrenched* in terms of its fundamental and unyielding nature, being positioned beyond the sway of law and politics. Yet, unlike constitutional provisions, or even more mundane legislation, architectural exclusion is lifted beyond public debate, scrutiny, and accountability. While emphasis should be placed upon "the importance of technical arrangements that precede the *use* of the things in question,"[78] Winner continues to observe that:

> To our accustomed way of thinking, technologies are seen as neutral tools that can be used well or poorly, for good, evil, or something in between. But we usually do not stop to inquire whether a given device might have been designed and built in such a way that it produces a set of consequences logically and temporally *prior* to any of its professed uses.[79]

The question then becomes what such prior consequences would be considered before the introduction of autonomous vehicles. Given the noted biases and disparate impacts observed for aligned technologies — even "just" big data[80] — it is difficult to imagine that autonomous vehicles could be designed and deployed in a socially neutral fashion. It should also be noted that, while there is substantiating evidence that Moses had intended for his architecture to be exclusionary, such intention (or even recklessness) is not necessary to bring about such consequences: "Rather, one must say that the technological deck has

(emphasis is added because the exclusionary barriers are placed to differentiate ease of access to that public park).

[77] *Id.* at 124 (emphasis in original). This view is neatly encapsulated by Lee Koppleman's statement: "The old son-of-a-gun had made sure that buses would *never* be able to use his goddamned parkways." *Id.* at 124 (quoting ROBERT A. CARO, THE POWER BROKER: ROBERT MOSES AND THE FALL OF NEW YORK 952 (1974) (emphasis in original)).

[78] *Id.* at 125 (emphasis in original).

[79] *Id.* (emphasis in original).

[80] *See generally* Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671 (2016) (showing that discriminatory tendencies exist even if they have not been manually programed, whether on purpose or by accident, because discrimination may be an artifact of the data mining process itself).

been stacked long in advance to favor certain social interests, and that some people are bound to receive a better hand than others."[81]

This line of argumentation suggests that the autonomous vehicle will usher in a new era replete with its own supporting infrastructure, which in turn will repeat previous versions of discrimination and segregation that have accompanied analogous developments in the past. The production of laws that dis-privilege pedestrians, for example, have a clear precedent in criminalizing jaywalking, unsurprisingly spurred by the growing automobile industry in the 1920s.[82] Not only were pedestrians "essentially written out of the equation when it came to designing streets," but "law-enforcement tend[s] to identify with a motorist's perspective,"[83] such that both urban design and law have historically conspired to segregate public space and to marginalize pedestrians.

Autonomous vehicles are likely to exacerbate the developments introduced by the automobile, since these are likely to adopt different strategies for confronting the challenges of navigating traffic than human drivers currently deploy. Ultimately, the efficiency and safety concerns that undergirded the introduction of jaywalking laws may justify the total exclusion of human beings from transport thoroughfares optimized for autonomous vehicles.[84] Excluding human beings would be a continuation of the privatization of public space,[85] this time under the guise of new technological capabilities. Such effects are likely to be spurred on by the need for cities to be competitive to raise the necessary funds to integrate autonomous vehicle infrastructure into their urban spaces.[86] As with Winner's observation of the longevity of Moses' public works in New York City, however, the manner in which this infrastructure for autonomous

---

[81] Winner, *supra* note 76 at 125-26.

[82] *See* Aidan Lewis, *Jaywalking: How the Car Industry Outlawed Crossing the Road*, BBC NEWS (Feb. 12, 2014), http://www.bbc.com/news/magazine-26073797.

[83] *Id.*

[84] For example, Copenhagen's fully-automated Metro functions reliably and effectively in part because human beings are excluded from its tracks.

[85] *See generally* ANNA MINTON, GROUND CONTROL: FEAR AND HAPPINESS IN THE TWENTY-FIRST-CENTURY CITY (2012) (arguing that the exclusion of people on properties built and owned by private corporations in British cities not only transformed these cities but also the nature of public space).

[86] In December 2015, the U.S. Department of Transportation launched "Smart City Challenge" to incentivize mid-sized cities to develop ideas for an integrated smart transportation system through national competition. *See Smart City Challenge*, U.S. DEP'T TRANSP., https://www.transportation.gov/smartcity (last visited Apr. 17, 2018).

vehicles is developed and the logics which it supports will dictate human behavior in the built environment long into the future.

Two particular problems arise with the prospect of further architectural exclusion arising from autonomous vehicle infrastructure. First, the streets are the paragon of public space,[87] and the cradle of freedom of speech and assembly essential to democratic debate and legitimate public protest.[88] Once the exclusionary features of the infrastructure are in place, it may be that such developments would be especially difficult to confront and resist because it removes the very means of assembly and protest that have traditionally precipitated change.[89] Reaching back towards the analogies between law and architecture as different modalities of regulation, curbing the ability to challenge and change the built environment may be similar to entrenchment of legal provisions.[90] Furthermore, there may be more subtle forms of discouragement at play:[91] for example those individuals without immunity devices or relatively high scores in the trump card hierarchy may avoid certain streets (or even avoid the streets altogether) for the increased risks displaced upon them. Crowds may even disperse during events where high status people are planned to congregate for similar reasons. This trajectory of logic would culminate in areas where members of the public are issued a "notice" to stay "off" the streets where autonomous vehicles are present — an extension of contemporary jaywalking laws — that may

[87] *See* ERIC BARENDT, FREEDOM OF SPEECH 273-90 (2d ed. 2007); *see also* JANE JACOBS, THE DEATH AND LIFE OF GREAT AMERICAN CITIES 29 (1961) ("Streets and their sidewalks, the main public places of a city, are its most vital organs.").

[88] *See* Frisby v. Schultz, 487 U.S. 474, 480 (1988) ("[P]ublic streets and sidewalks have been used for public assembly and debate, the hallmarks of a traditional public forum."); *see also* BARENDT, *supra* note 87 at 273-90.

[89] *See, e.g.*, *Frisby*, 487 U.S. at 484 (restricting the public's right to protest by upholding a law banning targeted picketing in residential streets due to the unique nature and privacy of homes). "One can point to Baron Haussman's broad Parisian thoroughfares, engineered at Louis Napoleon's direction to prevent any recurrence of street fighting of the kind that took place during the revolution of 1848." Winner, *supra* note 76, at 124. A crude comparison, but apt given how the layout of public streets can affect the ability and efficacy of protesting authority and power.

[90] *See generally* Hin-Yan Liu, *Constitutional Entrenchment: Questions of Legal Possibility and Moral Desirability in the United Kingdom*, 2 CITY UNIV. HONG KONG L. REV. 193 (2010) (showing that entrenchment is a legal protection of certain fundamental rights from being repealed by political powers and cannot be abrogated like ordinary laws).

[91] My thanks, again, to Lugar Sungil Choi for this suggestion.

then entrench themselves as legal and social practice analogous to how jaywalking laws have done since their introduction.[92]

Second, the parity of law and architecture as modalities of regulation suggest a degree of insulation between the influence of each upon the other. The point here is that when attempting to overturn architectural exclusion through the law, courts are unlikely to be as effectual as attempting to alter legislation through the built environment.[93] This point is neatly expressed by Schindler:

> Instead of garnering support to pass a law banning poor people or people of color from the places in which he did not want them — which, if the intent were clear, would not be permissible today — Moses used his power as an architect to make it physically difficult for certain individuals to reach the places from which he desired to exclude them.[94]

Any discriminatory effect that is incorporated into the infrastructure supporting autonomous vehicles will thus likely be difficult both to identify through legal lenses and to challenge in the courts of law. This makes this form of structural discrimination introduced by autonomous vehicles impervious to legal recognition, despite its disparate impacts.

Thus, technological progress, logistical expedience, safety concerns, urban design, and privatization may converge to entrench a transportation system dominated by autonomous vehicles that further oust the human from physical space and segregate classes of people according to their ability to access and use the services provided by such systems.

CONCLUDING THOUGHTS

A more nebulous insight arising from the debates surrounding the introduction of autonomous vehicles deploys that technology is a mirror to contemporary society and in particular its transport sector. That autonomous vehicles promise safer roads and offer a yardstick to lambast human drivers[95] might conversely suggest that the institution

---

92 *See supra* notes 84–87 and accompanying text.

93 *See* Schindler, *supra* note 42, at 1954.

94 *Id.* (citations omitted). It could also be noted that Moses' architectural approach to social exclusion was significantly longer-lived as a result of it utilizing an architectural, as opposed to legal, modality of regulation. *See supra* notes 79–81 and accompanying text.

95 *See* Alex Davies, *Uber's Latest Crash Proves We Need Self-Driving Cars*, WIRED (Mar. 25, 2017), https://www.wired.com/2017/03/uber-self-driving-crash-tempe-arizona/.

and practice of automobile driving is inherently a very dangerous activity. Rather than road accident and fatality statistics paving the way towards greater and fuller autonomy in vehicular systems, perhaps the reflection given to us by autonomous vehicles should undergird a re-evaluation of what a road "accident" actually means. The contemporary legal setting is such that accidents are framed under tort laws under the doctrines of negligence and product defects.[96] But if the proven and consistent dangers associated with driving become visible by technologies that promise to perform the same activities in a safer manner, might a more active legal stance be justified? In other words, the law might actively recognize the live and active danger inherent within driving as an activity, and the imposition of risk upon third-parties of undertaking such an activity based upon the defamiliarizing vantage point offered by autonomous vehicles. This might ground a re-evaluation as to whether the law should continue to treat traffic fatalities as "accidents" as marginal and unfortunate occurrences that are the by-products of efficient mobility and characterized by negligence and failure modalities of responsibility. Rather, the law should speak in more active terms of commission.

A predominant concern in writing this paper was to introduce a dystopian scenario[97] that may be capable of real-world implementation, thereby helping to bring it about. While certain aspects of systematic autonomous vehicle discrimination appear outlandish and implausible, when spread over several logical steps and implemented gradually, these effects may be rendered imperceptible. Other aspects, like the possibility of passive structural discrimination, appear to be corollaries of networked effects and will be inevitable, unless significant attention is paid to avoid the emergence of such outcomes.[98]

In order to mitigate the worst dimensions of these scenarios, the broadest range of participation in the design and development of these systems needs to be implemented. This is in order to offset the actor-orientation adopted by the ethical frameworks and to rebalance considerations to those who may have no access or involvement with the technology, regulation, or use of autonomous vehicles. This redress may also give voice to more marginalized individuals and groups within society who would otherwise be imposed greater

[96] *See generally* Bryant Walker Smith, *Automated Driving and Product Liability*, 2017 MICH. ST. L. REV. 1 (concluding that the current product liability regime is probably compatible with the adoption of automated vehicles).

[97] *See supra* Part III.

[98] *See supra* Part II.

burdens without consultation or acquiescence. Broad public engagement will also be necessary to establish impermissible types of development and will need to occur well in advance of technological maturation, and ideally at an earlier stage such that regulation can influence design and implementation. Finally, and more prosaically, soliciting a broad range of opinion could lessen the prospect for regulatory blind-spots and increase the confidence of both those driving the technology forward as well as potential users and victims of autonomous vehicles.

Direct countermeasures may need to be developed in order to decentralize coding, and to ensure that their operations do not coalesce around principles that are too similar. It may be worth considering imbuing autonomous vehicles to operate independently to some extent, or allow their human occupants to set their own preference settings and take responsibility for the outcomes arising from their directions.[99] Bright regulatory lines may be necessary to curb some of the excesses envisaged here, especially to prevent precisely the intentional discriminatory systems that might become profitably implemented.[100] It may be that crash-optimization as a framing consideration, being at root of the impulses to minimize total harm or maximize societal well-being, should be rendered irrelevant and impermissible. In this sense, uncertainty and chance will be intentionally implemented as more natural buffers against allegations of structured discrimination in order to disrupt patterned outcomes that might distill burdens and benefits to particular societal groups.

Beyond direct interventions, the dystopian scenario predicts the alignment of individual personal safety interests and commercial profit-driven incentives.[101] Regulation could be crafted to intervene and disrupt this nexus or impose various limits that decrease such drives. On the other end of the spectrum, existing codes of ethics for the engineers driving forward these technologies could be extended to cover possibilities of discrimination by countenancing the prospect for such emergent outcomes.[102] By starting this discussion of possible veiled systematic effects, the aim of this Article is to raise awareness of these undesirable possibilities and to foster a discussion to craft law and policy to mitigate at least the worst dimensions of these effects.

---

[99] *See, e.g.*, Contissa et al., *supra* note 15. For an opposing perspective, see Lin, *Adjustable Ethics Settings*, *supra* note 5.

[100] *See supra* Part III.

[101] *See supra* Part I.

[102] Lin, *The Robot Car*, *supra* note 5.

More difficult to counteract, at least through legal mechanisms, are the structurally discriminatory impact of architecture and the design of the built environment.[103] As law and architecture are set on par with each other as regulatory modalities, it is their combined effects that induce regulatory pressures.[104] But their subsidiary status as regulatory modalities also suggests that these have limited impact upon each other, such that exclusion or discrimination as effectuated through architecture will require different strategies for identifying and overcoming their effects.[105] Even still, impact assessments may be required to model the effects of the configuration of autonomous vehicle infrastructure before implementation as a bulwark against excessive effects.

This paper has largely treated these three forms of structural discrimination as distinct and independent. While each pose difficult problems individually, it is likely that the interaction effects will be the most pernicious and the most resilient against reform. The common denominator underlying structural discrimination is the crash-optimization imperative (and to some lesser extent, the demands of efficiency). A lazy approach would be to mitigate this imperative, for example by instilling randomness into the system. The underlying problem, however, is that autonomous vehicles are unlikely to remain *autonomous*, but instead will be coordinated in their behavior or will converge towards certain outcome patterns.[106] Thus, removing the overarching goal of crash-optimization will merely displace where the problems fall, rather than address the root cause at all.

[103] *See supra* Part V.

[104] *See supra* notes 73–77 and accompanying text.

[105] *See supra* note 95 and accompanying text.

[106] *See supra* notes 30–31 and accompanying text.