HOOVER IP²

WORKING PAPER SERIES
NO. 19002

AUTOMATORTS: HOW SHOULD ACCIDENT LAW ADAPT TO AUTONOMOUS VEHICLES? LESSONS FROM LAW AND ECONOMICS

ERIC TALLEY
COLUMBIA LAW SCHOOL

JANUARY 8, 2019

Hoover Institution Working Group on
Intellectual Property, Innovation, and Prosperity
Stanford University

www.hooverip2.org

Automatorts: How Should Accident Law Adapt to Autonomous Vehicles? Lessons from Law and Economics*

Eric Talley†

January 7, 2019

Abstract

The introduction of autonomous vehicles (AVs) onto the nation's motorways raises important questions about our legal system's adaptability to novel risks and incentive problems presented by such technology. A significant part of the challenge comes in understanding how to navigate the transition period, as AVs interact routinely with conventional human actors. This paper extends a familiar multilateral precaution framework from the law and economics literature by analyzing interactions between algorithmic and human decision makers. My analysis demonstrates that several familiar negligence-based rules (for precautions and product safety) are able to accommodate such interactions efficiently. That said, a smooth transition will likely require substantial doctrinal/legal reforms in certain states, as well as a more general reconceptualization of fault standards across all states—not only for AVs but also for for human actors themselves.

*PRELIMINARY AND INCOMPLETE DISCUSSION DRAFT. Do not cite or quote without permission. This paper grew out of joint work with Sharon Di, Qiang Du and Xi Chen, and was supported by a generous seed grant from the Columbia Data Science Institute. Thanks to each of these collaborators for their lively exchanges. Thanks as well to Jonathan Steinberg for excellent research assistance. All errors are mine.

†Isidor & Seville Sulzbacher Professor, Columbia Law School and Columbia Data Science Institute, Columbia University. Email: *etalley@law.columbia.edu.*

1 Introduction

In the spring of 2018, Tempe Arizona became the scene of a first-of-its-kind traffic fatality, when a Volvo XC90 owned by Uber and operating in autonomous mode struck a 49-year-old pedestrian. She died instantly.[1] Although the vehicle was not *completely* autonomous (it carried a human monitor whose conduct has itself attracted extensive scrutiny[2]), the incident elicited considerable alarm among policy makers, concerned not only about the evident shortcomings in the technology, but also about whether our legal and regulatory infrastructure is equipped to take on increasingly frequent interactions between humans and fully autonomous vehicles (AVs).

The legal and regulatory challenges are far from trivial, and they are perhaps all but inevitable: Under Arizona law (and that of many other states), legal liability for multi-person traffic accidents typically entails a complex calculus of comparative liability assessments between all participants. The involvement of an autonomous vehicle complicates matters further by adding other parties to the mix, such as designers of a AV software and hardware manufacturers.

Most commentators anticipate that product liability laws will come to play dominant role in these interactions, particularly as human drivers are pushed to the outer periphery of an increasingly networked driverless-car ecosystem (Anderson et al. 2014; Geistfeld 2017). While that endpoint may ultimately simplify the legal analysis in some ways (by removing some of the potentially contributing actors), it makes even more pressing the need to apportion liability risk among accident victims, AV owners/passengers, and the businesses who manufacture, design and market autonomous vehicles. And, the challenge is perhaps the greatest in the "transition" period we are now entering, as autonomous vehicles interact with a variety of still-*non-peripheral* humans (drivers, pedestrians and cyclists[3]). Such human actors will be increasingly exposed to self-driving vehicles alongside traditional drivers, espe-

[1] New York Times, "Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam" (By Daisuke Wakabayashi, March 19, 2018).

[2] The Verge, "Uber manager raised concerns about self-driving program just days before fatal collision" (By John Porter, December 11, 2018). Reuters, "Uber's use of fewer safety sensors prompts questions after Arizona crash" (By Heather Somerville, Paul Lienert, Alexandria Sage, March 27, 2018).

[3] It merits observing that even if human drivers were displaced *entirely* by AVs, AV/human interactions would continue indefinitely among cyclists and pedestrians.

cially in dense urban areas. It is plausible that human agents will attempt to "outsmart" self-driving technology, taking for granted that AVs will respond with effective countermeasures, thereby taking fewer precautions themselves. Here multiple challenges ensue, not the least of which is understanding what constitutes "reasonable" behavior by prospective victims, since it involves assessing the strategic interaction between humans and machines. For example, suppose a pedestrian identified a vehicle as self-driving, and on that basis began to behave more carelessly (e.g., assuming the autonomous vehicle would accommodate her erratic behavior). Is it more justifiable for her to expect that the algorithm could adapt, or that the technology might fail? Should our liability system countenance such behavior? Might there be circumstances where human actors should be held to a *higher* standard when interacting with self-driving cars? On the other hand, if the purveyors of self-driving technology anticipated tepid legal oversight, they may cut too many corners in designing navigation algorithms, imposing undesirable externalities on other participants. The interplay of these considerations is complex, both computationally and conceptually.

The fast moving nature of technological progress raises the stakes further by disrupting the informational environment. Today's self-driving cars stand out in a crowd, usually identified through the characteristic, 360-degree spinning LIDAR units on the roof (and often with no one in the driver seat). These obvious AV "tells" can be both a burden and a benefit for legal designers. On the one hand, as human beings become adept at recognizing driverless cars, their incentives to behave opportunistically may increase.[4] On the other, such clear markers make it possible for the legal regime to impose duties in a context-specific way, depending on the precise combination of human and technological actors at play. That said, as AV technology becomes increasingly embedded into (rather than onto) automobiles, the task of discerning an autonomous vehicle may become prohibitively demanding. Is that a good development?

This paper attempts to tackle some of these risk, incentive and information questions using tools from the economic analysis of law, extending a familiar analytical framework (e.g., Shavell 2007) to glean better understanding of how the liability system can best allocate the costs of accidents during the transition period; my specific aim is to ask what sorts of liability struc-

[4] See, e.g., New York TImes, "Wielding Rocks and Knives, Arizonans Attack Self-Driving Cars" (By Simon Romero, Dec. 31, 2018).

tures are best suited to a mixed ecosystem with both AVs and human actors operating simultaneously, and accordingly what challenges we are likely to face in the medium term. I argue that the introduction of AVs into a (heretofore) human-dominated ecosystem not only has implications for how we treat AVs (as many have correctly noted), but also how we regulate the behavior of human drivers. When viewed through the lens of the status quo ante, the tort system can be conceived as working to solve a "multi-sided" investment problem amongst human drivers themselves. Drawing on a canonical law-and-economics analytic framework (the "standard framework" e.g., Brown 1973), I conceive of accidents as the endogenous by-product of precautionary investments undertaken by individual actors—investments that jointly have the effect of reducing the expected likelihood of an accident in any given interaction. This setup is in many ways an instantiation of a team production problem (e.g., Holmstrom 1982), and accordingly it raises well-known collective action problems associated inducing optimal investments by participants. When precaution measures are verifiable *ex post*, however, a variety of familiar fault-based rules are (under certain assumptions) able to incentivize optimal care. My innovation here is to extend the standard framework to a setting where some (but only some) actors may decide *ex ante* to employ algorithmic AV technology to govern their driving before entering traffic. In so doing, AV adopters largely neuter their individual abilities to invest in precautions while driving, depending instead on up-front investments in product design. Although algorithmic decision making works flawlessly when functioning within the AV's design parameters, the breadth of that design is chosen endogenously (and at a cost) by the AV adopter.[5] And, when the algorithm confronts situations outside its pre-specified parameters (*i.e.,* in unfamiliar or unanticipated contexts), it becomes erratic and unpredictable.

Preliminary analysis of this framework yields several intuitive conclusions. First (and most obviously), it reveals that the predominant traditional means for regulating bilateral accidents with solely human actors—predicated on assessing in-the-moment precautions undertaken by injurers and victims—is maladapted to the hybrid environment. By definition, AV technology supplants in-the-moment human judgment, and it is thus virtually incoherent

[5] In the Uber incident in Arizona, for example, the AV was equipped with a single LIDAR device, as compared with the seven that its predecessor test vehicles used. This decision limited the vertical field of vision of the AV technology. Reuters, "Uber's use of fewer safety sensors prompts questions after Arizona crash" (By Heather Somerville, Paul Lienert, Alexandria Sage, March 27, 2018).

to ask whether the algorithm has exercised "due care" in taking precautions while driving. Second, notwithstanding the ill-fitting nature of bilateral precaution framework for AV/human interactions, reverting to a simple "no-fault" or "strict liability" regime is no panacea either: such regimes suffer from well-known deficiencies in providing multilateral incentives to avoid harm, and those deficiencies remain just as profound in the presence of endogenously designed algorithms. Third, because the only way to apply legal scrutiny to AV decision making is at the *ex ante* design/build stage, an optimal liability regime necessarily onboards certain attributes of product liability. Finally, I demonstrate that there are several product liability regimes which—when coupled with a due-care regime for human actors—are capable of inducing efficient behavior under the "transitional" ecosystem I focus on here: these include both strict liability subject to a contributory/comparative negligence defense, as well as ordinary negligence (also potentially subject to a contributory/comparative negligence defense).[6] And in fact, these two candidates in particular appear to match up well to the *status quo ante* among most states' existing product liability regimes.

That said, even if current institutions can accommodate the AV/human transition, there are several challenges that such institutions are likely to face. First, a small number of states currently have product liability doctrines that appear to steer close to a pure strict liability regime with no contributory/comparative negligence defense. Such a system tends (within the framework analyzed) to catalyze significant underinvestment in precautions by non-adopters of AV. Additionally, another (more sizable) portion of states similarly leans toward strict liability for product liability claims, but they clearly recognize contributory/comparative fault defenses. Although such a regime can bring about efficient investments in precaution / product design, it has far distinct distributive properties than the negligence rule that regulates human drivers: while an effective negligence regime tends to impose residual risk on the victim, a strict-liability regime (with defenses) tends to impose it on the injurer. Consequently, the decision to adopt AV would likely be distorted downward (and indeed over-deterred) by the prospect of having to take that residual risk on board. Finally, even for the majority of jurisdictions that effectively embrace a negligence rule for product liability, the introduction of AV is likely to have implications for the appropriate standard of care that is applied to human participants in AV-related accidents. In

[6] These terms are defined more precisely in Section 2 below.

fact, my analysis suggests that when prospective human victims are readily able to discern AVs from human drivers, an efficient doctrine might either *increase* or *decrease* the contributory negligence standard of care for human victims relative to human-human interactions, depending on the situational context that generates the accident.[7]

My analysis proceeds as follows. Section 2 provides a brief institutional overview of the status quo ante in accident law across US jurisdictions. Section 3 develops a theoretical framework of multi-lateral investment by injurers and victims fashioned after the standard account in the law and economics literature, extending the baseline model to consider the introduction of AV technology in a hybrid setting, where some AV adopting injurers interact with human victims. Section 4 discusses a variety of extensions to the model to explore robustness. Section 5 concludes.

2 Doctrinal Landscape

Before plunging into analytics, this section aims to provide a brief orientation for the uninitiated as to how American tort law apportions liability risk in vehicular accidents. The history is a long one, ably chronicled by acknowledged experts in the field. (*See, e.g.,* Engstrom 2018; Rabin 2005; Geistfeld 2017; and Mashaw & Harfst 1990. Interested readers should consult those sources for a richer set of details.) In short, however, it is fair to say that transportation accidents shaped large swaths of the tort law landscape we have inherited today. In delivering this brief overview, I will refer many times of Table 1 below,[8] which illustrates a state-by-state comparison of certain central features of the US tort law system as it pertains to automobile accidents. I subdivide discussion into regulation of driver precautions, and (of particular relevance to AV technology) product liability claims.

[7] And even outside of such complete information environments, the optimal comparative/contributory fault standard for human victims would still be appreciably different in a hybrid setting. See Section 3(b).

[8] This table was assembled from a variety of sources, including Kroll & Westerlind (2012); American Bar Association (2009); Maryland Department of Legislative Services (2004); and propriatary research. Contact the author for a more detailed table with citations.

	Vehicular Accidents			Products Liability			Puntive
State	Standard	Cont/Comp Negligence	Cutoff	Standard	Cont/Comp Negligence	Cutoff	Damages
AL	Negligence	Contributory	n/a	Negligence	Contributory	n/a	UnCapped
AK	Negligence	Pure Comparative	n/a	Strict Liability	Pure Comparative	n/a	Capped
AZ	Negligence	Pure Comparative	n/a	Strict Liability	Pure Comparative	n/a	UnCapped
AR	Negligence	Modified Comparative	<50%	Strict Liability	Modified Comparative	<50%	Capped
CA	Negligence	Pure Comparative	n/a	Strict Liability	Almost Pure Comparative	n/a	UnCapped
CO	Negligence	Modified Comparative	<50%	Strict Liability	Modified Comparative	<50%	Capped
CT	Negligence	Modified Comparative	≤50%	Strict Liability	Almost Pure Comparative	≤50%	Capped
DE	Negligence	Modified Comparative	≤50%	Negligence	Modified Comparative	≤50%	UnCapped
DC	Negligence	Contributory	n/a	Strict Liability	Contributory	n/a	UnCapped
FL	Negligence	Pure Comparative	n/a	Strict Liability	Pure Comparative	n/a	Capped
GA	Negligence	Modified Comparative	<50%	Strict Liability	Modified Comparative	<50%	Capped
HI	Negligence	Modified Comparative	≤50%	Strict Liability	Modified Comparative	≤50%	UnCapped
ID	Negligence	Modified Comparative	<50%	Strict Liability	Modified Comparative	<50%	Capped
IL	Negligence	Modified Comparative	≤50%	Strict Liability	Modified Comparative	≤50%	UnCapped
IN	Negligence	Modified Comparative	≤50%	Strict Liability	Modified Comparative	≤50%	Capped
IA	Negligence	Modified Comparative	≤50%	Strict Liability	Modified Comparative	≤50%	UnCapped
KS	Negligence	Modified Comparative	<50%	Strict Liability	Modified Comparative	<50%	Capped
KY	Negligence	Pure Comparative	n/a	Strict Liability	Contributory	n/a	UnCapped
LA	Negligence	Pure Comparative	n/a	Strict Liability	Pure Comparative	n/a	UnCapped
ME	Negligence	Modified Comparative	<50%	Strict Liability	Modified Comparative	<50%	Capped
MD	Negligence	Contributory	n/a	Strict Liability	Contributory	n/a	UnCapped
MA	Negligence	Modified Comparative	≤50%	Negligence	Modified Comparative	≤50%	UnCapped
MI	Negligence	Almost Pure Comparative	≤50%	Strict Liability	Almost Pure Comparative	≤50%	UnCapped
MN	Negligence	Modified Comparative	≤50%	Strict Liability	Modified Comparative	≤50%	UnCapped
MS	Negligence	Pure Comparative	n/a	Strict Liability	Pure Comparative	n/a	UnCapped
MO	Negligence	Pure Comparative	n/a	Strict Liability	Pure Comparative	n/a	UnCapped
MT	Negligence	Modified Comparative	≤50%	Strict Liability	Modified Comparative	≤50%	UnCapped
NE	Negligence	Modified Comparative	<50%	Strict Liability	Modified Comparative	<50%	None
NV	Negligence	Modified Comparative	≤50%	Strict Liability	Modified Comparative	≤50%	Capped
NH	Negligence	Modified Comparative	≤50%	Strict Liability	Modified Comparative	≤50%	None
NJ	Negligence	Modified Comparative	≤50%	Strict Liability	Modified Comparative	≤50%	UnCapped
NM	Negligence	Pure Comparative	n/a	Strict Liability	Pure Comparative	n/a	UnCapped
NY	Negligence	Pure Comparative	n/a	Strict Liability	Pure Comparative	n/a	UnCapped
NC	Negligence	Contributory	n/a	Negligence	Contributory	n/a	UnCapped
ND	Negligence	Modified Comparative	<50%	Strict Liability	Modified Comparative	<50%	Capped
OH	Negligence	Modified Comparative	≤50%	Strict Liability	Modified Comparative	≤50%	UnCapped
OK	Negligence	Modified Comparative	≤50%	Strict Liability	None	n/a	UnCapped
OR	Negligence	Modified Comparative	≤50%	Strict Liability	Modified Comparative	≤50%	UnCapped
PA	Negligence	Modified Comparative	≤50%	Strict Liability	None	n/a	UnCapped
RI	Negligence	Pure Comparative	n/a	Strict Liability	None	n/a	UnCapped
SC	Negligence	Modified Comparative	<50%	Strict Liability	None	n/a	UnCapped
SD	Negligence	Modified Comparative	<<50%	Strict Liability	None	n/a	UnCapped
TN	Negligence	Modified Comparative	<50%	Strict Liability	Modified Comparative	<50%	UnCapped
TX	Negligence	Modified Comparative	≤50%	Strict Liability	Modified Comparative	≤50%	Capped
UT	Negligence	Modified Comparative	<50%	Strict Liability	Modified Comparative	<50%	UnCapped
VT	Negligence	Modified Comparative	≤50%	Strict Liability	Modified Comparative	≤50%	UnCapped
VA	Negligence	Contributory	n/a	Strict Liability	Contributory	n/a	Capped
WA	Negligence	Pure Comparative	n/a	Strict Liability	Pure Comparative	n/a	None
WV	Negligence	Modified Comparative	<50%	Strict Liability	Modified Comparative	<50%	UnCapped
WI	Negligence	Modified Comparative	≤50%	Strict Liability	Modified Comparative	≤50%	UnCapped
WY	Negligence	Modified Comparative	≤50%	Strict Liability	None	n/a	UnCapped

Table 1: Summary of Vehicular Accident and Product Liability Laws in the U.S.

2.1 Driver Precautions

In regulating civil liability for vehicular accidents (other than deliberate acts of harm with a vehicle), US jurisdictions overwhelmingly employ some version of the "negligence" standard to judge the behavior of drivers. That is, drivers owe other actors on the highway system a duty to take reasonable precautions (or "due care") as they drive. As its name implies, the negligence standard is just that – a standard – and thus compliance tends to be measured on a situational basis, against the facts and circumstances surrounding the accident.

Although there are many ways to characterize "due care," a particularly useful one for this paper is to cast it in utilitarian, cost-benefit terms: Under this formulation, an incremental precaution is deemed required if, in the facts and circumstances then prevailing, the expected marginal benefit of taking the precaution (*e.g.,* through reduced accident probabilities and/or severities) exceeds the marginal cost of the precaution. Should a driver fail to take a precaution that a negligence standard would require, and should that omission causes a harm (both "in fact" and "proximately"), then the negligent driver is liable for all damages that an injured party(ies) can prove. On the other hand, if the harm could not have been avoided, or was avoidable only with precautions deemed "unreasonable" (pursuant to the same marginal cost-benefit test above), then the injured party cannot recover (even if it is clear that an accident would have been averted had the driver taken the unreasonable precaution). In other words, a negligence standard does not require one to take precautions that are unreasonably costly or inconvenient (relative to anticipated benefits) in order to avoid civil liability.

Even as US jurisdictions have overwhelmingly embraced the negligence standard for automobile accidents (see Table 1), it merits noting that this is not the only fault standard one might conceive. Under a strict liability standard, for instance, the injurer would be responsible for *all* harms caused by her failure to take precautions (regardless of how cost-effective such precautions would have been). Alternatively, under a recklessness standard, the driver is expected to take only those precautions that more-than-comfortably pass a cost benefit test – *i.e.,* whose expected marginal benefits are substantially larger then their marginal costs.

While negligence sounds simple enough on first blush, applying it gets complicated in many ways—one of them being the fact drivers are not merely navigating an obstacle course of inanimate objects. Rather, they are participating in multi-party ecosystem featuring many active decision makers. When an accident occurs, it frequently involves multiple agents (drivers/pedestrians/cyclists) who are simultaneously making precautionary choices (or–in some cases–fail to do so). Courts realized early on that vehicular harms were not infrequently the product of lapses by victims as well as injurers (and, perhaps, even third parties).

How does the tort system treat a negligent driver who causes harm to another party who was also negligent? Until the mid-20th century, most American jurisdictions embraced the doctrine of "contributory negligence," providing an affirmative defense to a negligent defendant who could show that the victim was also negligent herself, thereby avoiding liability altogether. To this day, a small handful of states (such as Alabama, Maryland and Virginia) continue to adhere to contributory negligence as an affirmative defense (See Table 1, Column 3). Over time, however, through judicial precedent and/or statutory reform, most jurisdictions began to move towards what is now commonly referred to as "comparative" negligence.[9] Comparative negligence yields the same result as contributory negligence when the injurer alone is found to be negligent: The injurer remains 100% liable. However, when the injured party is *also* judged to negligent, comparative negligence regime (in its "pure" form) requires the judicial fact finder to weigh the relative fault between the negligent plaintiff and the negligent defendant(s) (and sometimes even that of third parties). If the court determines that the plaintiff is (say) 40 percent responsible and the defendant is 60 percent responsible, then the plaintiff is allowed to collect only that 60 percent of his total damages allocated to the defendant.[10] There are a handful of states adhering to this "pure" form of comparative fault (including California and New York).

The remaining significant majority of states (including Illinois, Massachusetts and Texas) have embraced what amounts to an amalgam of comparative and contributory negligence rules, popularly known as "modified" comparative negligence. Like pure comparative negligence, this regime requires a court to assess and weigh the negligent parties' relative degrees of fault; but the modified rule also asks whether plaintiff's adjudicated share of

[9] Early judicial movements were in Florida and California. See Hoffman v. Jones, 280 So.2d 431 (Fla. 1973); Li v. Yellow Cab Co., 532 P.2d 1226, 13 Cal.3d 804 (1975).

[10] On yet another layer of complexity, if a court finds that some of the negligence responsibility should be assigned to non-parties in the litigation, then the defendant may also be held responsible for the share of non-parties too (depending on whether the jurisdiction allows for "joint and several" liability — a topic beyond the scope of this paper).

fault is strictly less than (or, in some cases no higher than) a specified critical cutoff (usually around 50% – See Table 1, Column 4). If the plaintiff's share satisfies the cutoff criterion, then the usual proportioning rules of pure comparative negligence apply. However, if the plaintiff's share is above the cutoff, then the rule functions as a contributory negligence standard, and all recovery is withheld.

2.2 Product Liability

As noted in Section 1 (and developed more below), the increasing penetration of driverless cars on American roadways is likely to magnify emphasis on the use of another branch of tort law—product liability—to assess liability when an accident involves an autonomous vehicle. Though broadly viewed as a part of tort law, this doctrine originally grew out of the law of contracts, and specifically long-standing doctrines pertaining to warranties of quality that are "implied" pursuant to a sale of goods and/or services.[11] Under these doctrines, the purchaser of a defective product could claim damages caused by any failure of the product to satisfy the implied minimal quality threshold. Although warranty claims can be powerful, they are a traditional form of contract right, and thus their availability is limited (with few exceptions) to parties who are in a direct contractual privity with one another. A harmed consumer's ability to enforce a warranty extends (at best) to the direct retailer of the good—with whom the consumer was presumably a direct contractual counterparty.[12] But if the retailer is not also the manufacturer, or if the harmed party is a third party and not the purchaser, then the liability trail goes cold.

Perhaps the most significant evolutionary moment in product liability law was its early 20th Century move to relax the privity requirement that warranty law usually requires. It is perhaps ironic justice that this evolutionary moment is acknowledged to have occurred in a dispute pertaining to an allegedly defective automobile: *MacPherson v. Buick Motor Co.*[13] (Mashaw & Harfst 1990). In *MacPherson*, a purchaser of an automobile was injured

[11] The chief forms of implied warranty are for "merchantability" and "fitness for a particular purpose." See Uniform Commercial Code Sections 2-314, 2-315.

[12] Note that even this warranty claim would be a stretch if, say, a third party other than the purchaser were injured due to a product defect. Here, too, the victim would have no direct contractual privity even with the retailer.

[13] MacPherson v. Buick Motor Co., 217 N.Y. 382, 111 N.E. 1050 (1916).

when a defective wheel collapsed on a car he had purchased from a dealer, who had in turn purchased it from the defendant manufacturer (Buick). In affirming a lower court's holding for the plaintiff, Justice Cardozo laid out the template of modern products liability law:

> If the nature of a thing is such that it is reasonably certain to place life and limb in peril when negligently made, it is then a thing of danger. Its nature gives warning of the consequence to be expected. If to the element of danger there is added knowledge that the thing will be used by persons other than the purchaser, and used without new tests, then, irrespective of contract, the manufacturer of this thing of danger is under a duty to make it carefully. That is as far as we need to go for the decision of this case....There must also be knowledge that in the usual course of events the danger will be shared by others than the buyer.... If [the manufacturer] is negligent, where danger is to be foreseen, a liability will follow.[14]

MacPherson's most celebrated innovation for products liability law was its break with the contractual privity requirement to trigger manufacturer liability in "warranty-like" claims. But the holding also appears—at least on the basis of its text—to ground the manufacturer's liability in a tort-like theory of *negligence*. This is an odd fit with traditional warranty law, which is largely the province of *strict liability*, wherein the warranting party's exposure is independent of her precaution efforts (or lack thereof). Thus, if one understands the case as simply extending warranty enforcement rights to foreseeable third parties (as some have construed it), strict liability should have followed—except that it didn't. In the wake of *MacPherson*, much confusion ensued.

And it substantially continues today. Indeed, the interpretive conundrum over what product liability actually *is* has never been resolved completely. As reflected in Table 1, the vast majority of US jurisdictions take a rhetorical cue from warranty law, describing their own product liability regime as imposing *strict liability* on defendants. In so doing, most of these state-level authorities work from a template established by the Restatement 2nd of Torts, Section 402A, which reads:

[14] 111 NE at 1053.

1. One who sells any product in a defective condition unreasonably dangerous to the user or consumer or to his property is subject to liability for physical harm thereby caused to the ultimate user or consumer, or to his property, if:

   (a) the seller is engaged in the business of selling such a product, and

   (b) it is expected to and does reach the user or consumer without substantial change in the condition in which it is sold.

2. The rule stated in Subsection (1) applies although

   (a) the seller has exercised all possible care in the preparation and sale of his product, and

   (b) the user or consumer has not bought the product from or entered into any contractual relation with the seller.

Significant ink has been spilled over the question of whether the text of Section 402A—and the myriad statutes / doctrines fashioned after it—truly announces a strict liability rule (as subsection 2(a) suggests), or instead smuggles in the contraband of negligence by conditioning liability on the product being in an "unreasonably dangerous" condition (per subsection 1). And important contours of product liability law in most states continue to reflect this jurisprudential schizophrenia, distinguishing (for instance) between "manufacturing" and "design" defects, with strict liability (or something close) attaching to the former and negligence (or something close) attaching to the latter.[15] In any event, it is now generally agreed that the "strict liability" label should not be read literally: it is a term of art that just as often signals a negligence standard for product manufacturers.

Accordingly, several challenges encumber the application of product liability law in the AV space. Primarily, the distinction between "design" and "manufacturing" defects is not always clear cut. Particularly when a defect relates to code or an algorithmic process, it arguably pertains to both the design of the product and its manufacture (Turner & Richardson 2002). And

[15] See Restatement 3rd of Torts, at 7-9 (reporter commentary to new Section 1). It is worth noting that the Third Restatement attempted valiently to "clarify" these distinctions, but did so constrained by a series of comprpromises with established state law doctrine, rendering a result that was (at least to this reader) is somewhat less than clear.

second, even if one could properly relegate AV defects to the design category, some states appear to take more literally the "strict-liability" labeling of product defects, at least when it pertains to contributory / comparative negligence defenses. Indeed, several states have simply refused apply their own general contributory/comparative negligence rules to product claims, taking the position that such defenses are simply anathema to a right that purports to be grounded in strict liability. (Examples of such include Oklahoma, Pennsylvania, Rhode Island and South Carolina – See Column 6 in Table 1). Put simply, product liability has long been in something of a state of flux when confronted with new technologies (Villasensor 2014). While that flux clearly imposes costs, it may also raise opportunities for tailoring a *sui generis* product liability doctrine for AV/human interactions. (I return to this issue in the next section.)

2.3 Punitive Damages

Finally, although not directly related to driver or product liability per se, I include a summary of punitive damages liability at the end of Table 1. Punitives are an assessment of monetary damages that are unrelated to compensating the injured party, but are instead awarded to punish the defendant for particularly wanton, malicious, cruel, or otherwise bad faith conduct.[16] Most states allow for punitive damages in torts cases, but some exclude either personal injury or wrongful death actions from eligibility. In addition, several states have imposed a cap on the maximal value of punitive damages available.[17] Beyond these explicit caps, the United States Supreme Court has held that constitutional due process concerns also constrain punitive damages, and that such awards become constitutionally suspect when they grow to an order of magnitude larger than compensatory damages.[18]

Although this overview has been a brief one, it provides a bit of the institutional back-story for the more analytical section that follows. In particular,

[16] In New York, for example, punitives are permitted when a defendant acts "recklessly, wantonly, or without regard to the rights of the plaintiff or of people in general." Hall v. Consolidated Edison Corp., 428 N.Y.S.2d 837 (1980). In California, punitives are available in cases involving oppression, fraud, or malice. Cal. Civil Code § 3294.

[17] In Texas, for example, exemplary damages may not exceed the greater of either $200,000 or two times the amount of economic damages, plus an amount equal to any non- economic damages found by the jury, not to exceed $750,000. Tex. Civ. Prac. & Remedies Code § 41.008.

[18] State Farm Mutual Automobile Insurance Co. v. Campbell, 538 U.S. 408 (2003).

it is worth noting that transitioning legal rules to a world of AV/human interaction will necessitate understanding how the two sides of Table 1 are likely to interact.

3 Framework and Model

This section of the paper uses a two-sided precaution model from the law and economics of torts to study the effects of introducing an autonomous vehicle (AV) technology. The framework is based on a long-established literature in law and economics. (See, e.g., Cooter & Ulen 2012; Shavell 2007; 1987; Brown 1973).

Two representative actors, denoted v (for "victim") and i (for "injurer") are assumed to interact in a context that might result in a vehicular accident. To concentrate on the most interesting application for autonomous vehicles, suppose that the injurer is *always* in a vehicle, while the victim could be a driver, a pedestrian, a cyclist, etc. In the event that the victim-injurer interaction results in an accident, the victim suffers a harm with expected value that is commonly-known to be $H > 0$. Whether an accident actually occurs hinges critically on the level of precaution undertaken by both parties. Let $x_v \geq 0$ and $x_i \geq 0$ denote (scalar) measures of precautions undertaken by the victim and injurer, respectively. Assume (for now) that these precautions are ex post verifiable, though they are chosen simultaneously. Given $\{x_i, x_v\}$, the probability that a harm occurs is given by $p(x_i, x_v)$, which is assumed strictly decreasing in both its arguments and convex. I also will frequently impose the assumption that the parties' precautions are weak substitutes for one another,[19] and that $\lim_{x_i \to \infty} p(x_i, x_v) = \lim_{x_v \to \infty} p(x_i, x_v) = 0$—so that accidents are (in principle) completely avoidable if either party is willing to exercise unbounded amounts of care.

In the following subsections,the analysis follows conventional law and economics approaches for accident law, assuming that both injurer and victim are natural persons who bear costs associated with precautions. In particular, in the each party incurs total cost of $w_i x_i$ and $w_v x_v$, reflecting constant marginal precaution costs of $w_i > 0$ and $w_v > 0$ for the injurer and victim (respectively). Although marginal costs are strictly positive, they are assumed sufficiently mild that it would be cost effective for either party to

[19] Mathematically, this is equivalent to assuming $\frac{\partial^2 p}{\partial x_i \partial x_i} \leq 0$ over the relevant range. I will be explict whenever this assumption is relaxed.

incur at least some precaution if the other party incurs none. (Because of the abstract functional form of $p(.)$, the constant marginal cost assumption is without loss of generality.)

After fleshing out the model in the context of human actors, I then extend the framework to study the introduction of autonomous vehicle technology with some fraction of injurers. (As will become apparent below, AV technology significantly changes the cost-of-precaution analysis from the injurer's perspective.)

3.1 Baseline Model: Human-Only Actors

Consider first the case (studied extensively in the literature) in which both injurer and victim are human actors.[20] To frame the analysis of legal institutions, it helps to establish a benchmark of an omniscient "Social Planner's" problem of fixing precautions to minimize expected social costs. Given the above parameters, total social costs associated with the parties' precautions are:

$$\Phi(x_i, x_v) = p(x_i, x_v) \cdot H + w_i \cdot x_i + w_v \cdot x_v \tag{1}$$

The social planner's problem is therefore to choose precaution levels to minimize social costs of precautions and harm:

$$\min_{(x_i, x_v) \geq 0} \Phi(x_i, x_v) \tag{2}$$

Let $\{x_i^*, x_v^*\}$ denote the first-best precaution levels for the two parties satisfying the above conditions, and assume hereafter that the solution to this problem is interior. The first order conditions associated with first-best bilateral precaution are:

$$\begin{aligned} H \cdot \frac{\partial p}{\partial x_i} + w_i &= 0; \\ H \cdot \frac{\partial p}{\partial x_v} + w_v &= 0 \end{aligned} \tag{3}$$

The intuition behind these conditions is simple – they state that at an interior optimum, the first-best raises each party's precautions to the point where the marginal benefit of precaution (taking the form of reduced expected harm

[20] Human in the rational, economic sense, that is – they still act to maximize expected equilibrium payoffs, and must be incentivized to expend precautions when interacting.

$H \cdot \frac{\partial p}{\partial x_i}$ and $H \cdot \frac{\partial p}{\partial x_v}$) is just offset by the respective marginal cost of investment (w_i and w_v). It is easily verified that x_i^* and x_v^* are strictly decreasing in w_i and w_v (respectively), and both are strictly increasing in H.

3.1.1 Efficient Legal Rules

The Law and Economics literature on accidents has devoted considerable attention to frameworks similar to baseline model in order to understand when (and under what circumstances) various civil liability regimes tend to induce first-best investments. It turns out that there are several candidates that are up to the task; they tend to be non-unique, but all tend to share characteristics consistent with "fault" (a.k.a., negligence) regimes. An intuitive place to start the analysis is to consider several familiar candidate legal regimes that–by contrast to negligence–*fail* to implement first-best.

No Fault Rule (NF) Consider first what is perhaps the simplest rule of all: no-fault (NF), whereby the victim must bear all the costs she incurs from accident, regardless of precautions. Because the cost of a harm to the injurer comes about only through the liability system, the injurer's and victim's expected *private* costs under a no-fault system are[21]:

$$\begin{aligned} \Phi_i(x_i|NF;x_v) &= w_i \cdot x_i \\ \Phi_v(x_v|NF;x_i) &= p(x_i,x_v) \cdot H + w_v \cdot x_v \end{aligned} \tag{4}$$

It is clear that regardless of the victim's precautions, it is strictly dominant under NF for the injurer to set $x_i = 0$, since she gains no benefit from precaution (e.g., in the form of avoiding liability). Anticipating this behavior, the victim will realize that she must internalize the entire social cost, and she will choose x_i to minimize:

$$\begin{aligned} \Phi_v(x_v|NF;0) &= p(0,x_v) \cdot H + w_v \cdot x_v \\ &= \hat{p}(x_v) \cdot H + w_v \cdot x_v \end{aligned} \tag{5}$$

where $\hat{p}(x_v)$ denotes $p(0,x_v)$. Here, conditional on no investment by the injurer, the victim will adopt precaution according to the condition:

$$\frac{\partial \hat{p}}{\partial x_v} \cdot H + w_v = 0 \tag{6}$$

[21]Note that these private costs sum to total social costs; i.e., $\Phi_i(x_i|NF;x_v) + \Phi_v(x_v|NF;x_i) = \Phi(x_i,x_v)$.

which also yields a unique solution of $\hat{x}_v$. Note that the victim's best-response correspondence in (6) is *functionally* identical to the first-best rule for the victim (3). This makes intuitive sense, since the victim bears the entire risk under a no-fault rule and thus internalizes all social costs and benefits on the margin. Nevertheless, because the injurer fixes precautions at $x_i = 0$ under NF, the overall level of care is clearly suboptimal relative to the first best, and the victim expends more precautions (at a greater private cost) then he otherwise would under the Social Planner's problem.[22]

Strict Liability Rule (SL) Now suppose that a strict liability rule governed the parties' behavior, such that the injurer is always liable for the victim's harm regardless of fault. In this case, the metaphorical tables are turned: for the victim now realizes that the legal rule grants her perfect insurance, and thus she bears no cost associated with harm. Consequently, the victim's private cost function is given by:

$$\Phi_v\left(x_v|SL; x_i\right) = w_v \cdot x_v \tag{7}$$

As with the injurer in the NF case, here the victim will have no incentives to take precautions, setting $x_v = 0$. Anticipating this choice, the injurer will choose precautions with the realization that he is the sole risk bearer:

$$\begin{aligned} \Phi_i\left(x_i|SL; x_v = 0\right) &= p\left(x_i, 0\right) \cdot H + w_i \cdot x_i \\ &= \tilde{p}\left(x_i\right) \cdot H + w_i \cdot x_i \end{aligned} \tag{8}$$

where $\overline{p}\left(x_i\right) \equiv p\left(x_i, 0\right)$. The injurer's first order condition is thus:

$$\frac{\partial \overline{p}}{\partial x_i} + w_i = 0 \tag{9}$$

yielding a unique optimum of $\overline{x}_i$, which (under the assumptions above) is strictly greater than the injurer's first-best level of precaution. Much like the NF case for the victim, the injurer adopts socially optimal *rule* for precautions, but it yields an outcome that remains suboptimal (because the victim invests nothing).

[22] This last point turns on the victim's and injurer's investments being structural substitutes in $p(.)$ as assumed above. When precautions are complements over the relevant margin, it is possible that $\hat{x}_i < x_i^*$.

Negligence Rules (N) Finally, consider a variety of rules emanating from the negligence "family". This group includes ordinary negligence, strict liability with a contributory negligence defense, ordinary negligence with a contributory negligence defense, and potentially several others.[23] Unlike NF and SL, these rules all commonly introduce a *discontinuity* in the underlying action space–a feature that (at least within the baseline set of assumptions) can incentivize efficient behavior by both sides. And as it turns out, numerous permutations within the negligence family can in fact induce first-best precautions, and are thus in this sense equivalent.[24]

Perhaps the simplest rule within the negligence family is an ordinary negligence rule, the injurer is held liable for the victim's injury only if the injurers level of precaution falls below a pre-specified fault standard—denoted x_i^N. Under such a rule, the injurer's expected payoff becomes explicitly discontinuous:

$$\Phi_i\left(x_i|N;x_v\right)=\begin{cases} p\left(x_i,x_v\right)\cdot H+w_i\cdot x_i & \Leftrightarrow x_i<x_i^N \\ w_i\cdot x_i & \Leftrightarrow x_i\geq x_i^N \end{cases} \tag{10}$$

The victim's expected cost is also discontinuous in x_i (at x_i^N):

$$\Phi_i\left(x_v|N;x_i\right)=\begin{cases} w_v\cdot x_v & \Leftrightarrow x_i<x_i^N \\ p\left(x_i,x_v\right)\cdot H+w_v\cdot x_v & \Leftrightarrow x_i\geq x_i^N \end{cases} \tag{11}$$

So long as the negligence standard is set at the injurer's first-best level—so that $x_i^N=x_i^*$—a simple negligence rule can induce first-best investment by both parties in equilibrium. To see why, suppose the victim invested optimally in precautions (at $x_v=x_v^*$) and consider the injurer's cost-minimizing choice. The first derivative of the injurer's expected cost is:

$$\Phi_i'\left(x_i|N;x_v\right)=\begin{cases} H\cdot\left.\frac{\partial p}{\partial x_i}\right|_{x_v=x_v^*}+w_i & \Leftrightarrow x_i<x_i^N \\ w_i & \Leftrightarrow x_i\geq x_i^N \end{cases} \tag{12}$$

Over the region where $x_i\geq x_i^N$, $\frac{\partial\Phi_i(x_i|N;x_v)}{\partial x_i}=w_i>0$, and the injurer wishes to choose the lowest possible precaution, setting $x_i=x_i^N$. In contrast, over

[23] There are several other permutations that work here as well, including (depending on its formulation) various comparative negligence regimes. I suppress them here to focus on intuitions, since on the equilibrium path comparative negligence replicates many of the featurs of contributory negligence (Shavell 2007).

[24] It warrants noting that they are *not* necessarily equivalent if one adds activity levels into the mix, as I detail in a subsequent section.

the region where $x_i < x_i^N$, the injurer's marginal cost is $H \cdot \left.\frac{\partial p}{\partial x_i}\right|_{x_v = x_v^*} + w_i$ which (by construction) is less than zero for all $x_i < x_i^N = x_v^*$, reaching a minimum as $x_i \to x_i^N = x_v^*$. Thus, assuming $x_i^N = x_v^*$, it is uniquely optimal for the injurer to invest $x_i = x_i^*$. From the victim's perspective the argument is similar: if the victim knows that the injurer will invest $x_i = x_i^N = x_i^*$, then his objective function becomes:

$$\Phi_i \left(x_v | N; x_i\right) = p\left(x_i^*, x_v\right) \cdot H + w_v \cdot x_v \tag{13}$$

His optimal precaution is characterized by the condition:

$$H \cdot \left.\frac{\partial p}{\partial x_v}\right|_{x_i = x_i^*} + w_v = 0, \tag{14}$$

which is identical to the first order condition stated in (3) for the first best optimum, so that the victim's optimal care choice sets $x_v = x_v^*$. It is clear that $\{x_i^*, x_v^*\}$ must be an equilibrium of the investment game when the negligence standard is set at the first-best precaution level for the injurer.

As noted above, simple negligence is not the only means by which a legal rule can implement first best in equilibrium. Several candidates within the negligence family can do the same. Consider, for example, a strict liability rule, but one subject to a contributory negligence defense: if the victim fails to exercise precaution of at least x_v^{CN}, then the victim's own negligence exonerates the injurer from bearing liability. Given the rough symmetry between the NF and SL cases above, it is perhaps easy to see that many of the arguments used to analyze negligence apply here too, and that fixing the contributory fault threshold at the first best (i.e., setting $x_v^{CN} = x_v^*$) can induce first best behavior in equilibrium.

A nearly identical set of arguments would also apply to other permutations of negligence regimes, including "nested" rules such as a negligence standard for the injurer subject to an affirmative defense of contributory negligence by the victim. There too, if the fault standard for the injurer is set at first-best, it remains an equilibrium for both parties to expend first-best level of precautions.[25] Similarly, it turns out that most plausible representations

[25] Note that when the negligence standard for the injurer is set optimally, in equilibrium the victim takes socially efficient precautions *regardless* of where the contributory fault standard is set. This is because compliance by the injurer with the negligence absolves implies that the victim becomes the sole risk bearer (and invests optimally).

of comparative negligence rules (both "pure" or "modified" comparative negligence rules, per Table 1) tend to lead to virtually identical analyses, and can be lumped into the "family" of negligence rules for current analysis.[26]

Several extensions to the basic model have also been explored, including harm visited on both parties when an accident occurs, risk aversion, measurement error as to precaution, uncertainty about the location of the negligence standard, and adding in activity levels. Although I cabin these extensions for now, it bears noting that the introduction of autonomous technology may affect many of these dimensions as well. (I return to these topics in a later section of this paper.)

3.2 Mixed Company: Algorithmic AV and "Pathways to Harm"

Having reproduced several well-known results from the law and economics of torts involving human actors, consider now how the introduction of autonomous vehicle (AV) technology alters (and disrupts) the incentive / risk landscape developed above. An important aspect of introducing AV technology is that human judgment at the time of an accident is supplanted by an algorithm that features (effectively) *preprogrammed responses to anticipated accident scenarios*, designed to avoid harm. At least when the algorithm encounters a scenario anticipated by the AV designer, I assume it execute a predetermined course of action that completely obviates an accident. Within such fully anticipated contexts, AV has an important leg up on human action, as the marginal cost of taking precaution for the autonomous actor effectively shrinks to zero.

That said, an important limitation of autonomous technologies (at least as conceived here) is that they operate in a deductive, rule-like fashion—one poorly adapted to novel or unfamiliar settings. Unlike human actors, whose intuitions about precautions are better able to transcend contexts (even to unfamiliar ones), algorithms tend to behave in rigid / unpredictable

[26] There are a few more strings attached here, since a comparative negligence rule further requires one to posit a mechanism for assessing relative fault. State doctrines or statutes tend to be silent about how to do so, and there are many ways to conceptualize the concept. Nevertheless, most plausible plausible assessment metrics tend to do the trick. (A common one is to presume that if both parties are deemed negligent—so that $x_v < x_v^*$; and $x_i < x_i^*$—the victim cannot recover the share of the total shortfall in precaution due to her: $\frac{(x_v^* - x_v)}{(x_v^* - x_v) + \left(x_i^* - x_i\right)}$.

manners when confronting unfamiliar decision making settings. To capture this intuition, I advance the conceptual develop the idea that each interaction of victim and injurer is set within a unique factual context that may result in an accident—each of which I shall refer to below—somewhat abstractly—as a *pathway to harm*. While a possibly infinite number of distinct pathways exist, the algorithm's design effectively determines the portion of them that the automated vehicle is capable of contending with.[27] Viewed from this perspective, then, the disruption to the system visited by AV technology is to substitute (1) pre-programmed responses to a pre-specified domain of pathways, specified (at a cost) *ex ante;* for (2) judgments that human actors are able to make adaptively in the moment (also at a cost of precaution, but transcending all pathways to harm).

To explore how this extension affects the analysis, this subsection introduces the possibility of AV technology into the standard bilateral-precaution model.[28] Within the modified framework, the victim faces exactly the same precaution technology as before, bearing constant marginal cost w_v to produce precaution at level x_v which, in concert with the injurer's precautionary intensity, results in an accident with probability $p(x_i, x_v)$. Per the discussion above, the victim's ability to take harm-reducing precaution is general and adaptable, transcending the specific "pathway" of harm.

As to the injurer population, I assume that some fraction $\alpha \in (0,1)$ of prospective injurers have adopted AV technology, delegating to a decision-making algorithm programmed to respond to a pre-specified domain of pathways.[29] Should the AV-injurer's situational interaction with the human victim fall within this domain, the injurer's marginal cost of harm avoidance is $w_i = 0$; *i.e.,* the algorithm can exercise all needed precautions (at zero marginal cost) to avoid an accident. In contrast, when the algorithm confronts a pathway that outside its prescribed domain, it behaves in a manner that is observationally equivalent to random behavior, contributing nothing to accident avoidance.[30] In such contexts outside the domain of the AV algorithm,

[27] It should be noted that even AV technologies that purport to be capable of some type of "deep learning" through experience tend to be poorly suited to completely unfamiliar settings.

[28] Recall that because the focus of this paper is on mixed systems where automated interact with human agents, I introduce AV technology at the injurer level (though a largely symmetric analysis is possible by introducing AV at the victim level).

[29] The remaining $1 - \alpha$ fraction of injurers are assumed to be human drivers, just as above. I take α to be exogenous for now, exploring the effects of endogenous uptake later.

[30] This formulation is somewhat stark, and it could probably be weakend somewhat;

then, it is as if the injurer's cost of precautions are infinitely large ($w_i \to \infty$).

Within this setting, much of the AV injurer's ability to contribute to harm avoidance stems from up-front engineering/design investments that expand the algorithm's domain,[31] rendering the AV capable of anticipating a larger and more robust set of pathways to harm. To capture this engineering problem formally, I suppose that for any victim-injurer interaction there is a continuum of mutually exclusive pathways to harm that govern the interaction between the injurer and victim. Let η denote a representative pathway, and assume that η is drawn from a set of mass one corresponding to the unit interval and distributed uniformly,[32] so that $\eta \tilde{} U[0,1]$. The realized value of η cannot be predicted *ex ante*, and nature randomly selects one to govern the the injurer and victim just prior to their interaction.

The injurer's AV design problem boils down to determining the domain of pathways that are to be "engineered into" the algorithm. Specifically, let $\gamma_i \in [0,1]$ denote the extent of the AV's capabilities (a manifestation of quality), which I assume to be a choice variable of the injurer. An AV technology of quality γ_i is capable of anticipating all pathways in the interval $[0, \gamma_i] \subseteq [0,1]$; for the subset of pathways defined by γ_i, the algorithm is able to avoid accidents entirely (functionally the same as driving x_i to ∞). The remaining pathways $\eta \in (\gamma_i, 1]$, in contrast, are beyond the algorithm's capacity to anticipate and confront (effectively driving $w_i \to \infty$ and $x_i \to 0$). Installing quality is costly, and the injurer must bear costs $c(\gamma_i)$ to install protection level γ_i, where $c(0) = 0, c'(\gamma) > 0; c''(\gamma) > 0$; $\lim_{\gamma \to 0^+} c'(\gamma) = 0; \lim_{\gamma \to 1} c'(\gamma) = \infty$.

The modified sequence is now as follows:

- Time 0: Liability regime set
- Time 1: Injurer/designer chooses γ_i at cost $c(\gamma_i)$.
- Time 2: Nature randomly selects whether the injurer is an AV adopter

at the same time, it captures an intuition about AV technology that many would concur with.

[31] The analysis assumes for simplicity that the AV design is conducted by the injurer, though in practical applications this would be by a third party connected by contract to the injurer. If there are negligible contracting / transaction costs between the AV designer and the injurer, then merging the two for the sake of current analysis sacrifices litte.

[32] This assumption is more restrictive than is necessary – similar results are obtained if η is distributed according to any cdf $F(.)$ with monotone hazard rate.

or human driver, with probabilities $(\alpha, 1-\alpha)$. In addition, nature chooses a pathway $\eta \in [0,1]$ governing the interaction of the parties.

- Time 3: Injurer and Victim implement precautions $\{x_i, x_v\}$ simultaneously:
  - Victim's precaution decision is as specified in the previous subsection, with victim incurring marginal cost w_v.
  - If a human injurer is selected, its precaution decision is also as specified above, with injurer incurring marginal cost w_i.
  - If an AV-adopting injurer is selected, its precautions are predetermined as a by-product of γ_i and the realized value of η :
    * If $\eta \leq \gamma_i$, injurer faces zero marginal costs of precaution and costlessly selects $x_i \to \infty$.
    * If $\eta > \gamma_i$, injurer faces infinite marginal costs of precaution and effectively selects $x_i = 0$.
- Time 4: Harm occurs (as before) with probability $p(x_i, x_v)$. Recall that $\lim_{x_i \to \infty} p(x_i, x_v) = 0$; in addition, recall that $\hat{p}(x_v) \equiv p(0, x_v)$.

A final dimension of the parties' interaction deserving attention (and motivated by the discussion in the introduction) is *information structure.* In particular, one must minimally specify the parties' awareness about (a) whether the injurer is an AV adopter or human driver; (b) if the injurer is an AV-adopter, what pathway realization (η) has obtained, and (c) what level of quality γ_i the AV injurer has installed. It is fair to assume that the injurer knows / can infer all of these facts. However, it is plausible that the victim might know (a) (b) and (c), only (a) and (b), only (a), or none of the above. In what follows, I suppose first that there is complete information, so that the victim can observe (a), (b) and (c); I then discuss the implications of relaxing these assumptions afterward.

3.2.1 AV Characteristics Observable to Victims

Consider the most straightforward scenario where the victim and injurer have symmetric information, so that the victim observes whether the potential injurer is utilizing AV; what quality level (γ_i) the injurer has installed, and the realized pathway η. (Assume further that these facts are verifiable to judicial actors.) In this situation, it is possible to condition both the social optimum and the legal rule on all information available to the parties at the moment of their interaction, effectively allowing human-human interactions to be treated separately from human-AV interactions. This separability implies, in turn, that the analysis is identical as in the previous section when victims interact with the $(1-\alpha)$ fraction of injurers who have not adopted AV; and—just as before—a variety of legal regimes from the negligence family are capable of supporting first-best precautions in equilibrium.

As to the remaining α fraction of the injurer population adopting AV, the first step is to characterize the socially optimal benchmark, constructed as follows. Suppose γ_i is fixed, and consider two potential realizations of η :

- $\eta \leq \gamma_i \Rightarrow x_i \to \infty$. In this scenario, total interim social costs are equal to:

$$\begin{aligned} \Phi\left(x_i, x_v | \eta, \gamma_i\right) &= \lim_{x_i \to \infty} p\left(x_i, x_v\right) \cdot H + w_v \cdot x_v \\ &= w_v \cdot x_v, \end{aligned} \tag{15}$$

  where the second equality is due to the fact that $x_i \to \infty$ in an anticipated pathway and $\lim_{x_i \to \infty} p\left(x_i, x_v\right) = 0$. It is clear that it is socially optimal for the victim to incur no precautions under this scenario, since such efforts would merely be duplicative of the algorithm's abilities. And thus, for $\eta \leq \gamma_i$, the socially optimal victim precaution is $x_v = 0$, and the maximized interim level of total social costs is $\lim_{x_i \to \infty} \Phi\left(x_i, 0 | \eta, \gamma_i\right) = 0$. This is not only first best, but it dominates what is attainable at all when AV is not available.

- $\eta > \gamma_i \Rightarrow x_i = 0$. In this case, interim social costs are given by:

$$\begin{aligned} \Phi\left(0, x_v | \eta, \gamma_i\right) &= p\left(0, x_v\right) \cdot H - w_v \cdot x_v \\ &= \hat{p}\left(x_v\right) \cdot H - w_v \cdot x_v \end{aligned} \tag{16}$$

  The first-best precaution for the victim is equal to $\hat{x}_v$, whose value is characterized by the optimality condition:

$$\frac{\partial \hat{p}}{\partial x_v} \cdot H = -w_v, \tag{17}$$

  yielding total social costs of $\Phi\left(0, \hat{x}_v\right)$. Note that $\Phi\left(0, \hat{x}_v | \eta, \gamma_i\right) > \Phi\left(x_i^*, x_v^*\right) > 0$, and thus the interaction of an AV injurer and a human victim can yield a *less* desirable interim outcome than with two human actors whenever $\eta > \gamma_i$, since the algorithm's capacity limitations are exceeded and all precautions must be borne by the victim. (Indeed, note that the outcome here is identical to the no-fault regime in human-human interactions discussed above.)

Backing up to the ex ante investment stage, and assuming optimal precautions (per the discussion above) are implemented downstream, social costs as measured ex ante are as follows:

$$\begin{aligned} \Omega(\gamma_i|\hat{x}_v) &= c(\gamma_i) + \gamma_i \cdot \Phi(\infty, 0|\eta, \gamma_i) + (1-\gamma_i) \cdot \Phi(0, \hat{x}_v|\eta, \gamma_i) \\ &= c(\gamma_i) + (1-\gamma_i) \cdot \Phi(0, \hat{x}_v|\eta, \gamma_i) \end{aligned} \quad (18)$$

The associated first order condition for minimizing social costs is thus:

$$c'(\gamma_i) = \Phi(0, \hat{x}_v|\eta, \gamma_i) \quad (19)$$

Let $\hat{\gamma}_i$ denote the value of γ_i satisfying this expression. The interpretation of this condition is that the injurer should continue to invest in expanding the AV's pathway domain until the marginal cost of quality improvement $(c'(\gamma_i))$ is just offset by the marginal social cost avoided $(\Phi(0, \hat{x}_v|\eta, \gamma_i))$ at the outer threshold of the AV's capacities.

Summarizing, the first-best optimum associated with and AV injurer / human victim interaction with symmetric information is:

$$\gamma_i = \hat{\gamma}_i;\ x_i = \begin{cases} \infty & \Leftrightarrow \eta \leq \hat{\gamma}_i \\ 0 & \Leftrightarrow \eta > \hat{\gamma}_i \end{cases};\ x_v = \begin{cases} 0 & \Leftrightarrow \eta \leq \hat{\gamma}_i \\ \hat{x}_v & \Leftrightarrow \eta > \hat{\gamma}_i \end{cases} \quad (20)$$

There are a few characteristics worth noting in comparing this scenario to its human-only counterpart. First (and unsurprisingly), the socially optimal actions now turn critically on harm realization of the pathway (η), while in the human-human case this pathway is irrelevant since human precautions are trans-contextual. Second, when the realized pathway falls within the AV technology's programmed parameters, it is both easy and optimal for the human actor to abandon precautions—indeed, this is one of the potential external benefits of introducing AV. Third, when the realized pathway falls outside the AV technology's domain, the prescribed level of precaution for the victim $(\hat{x}_v)$ will tend to differ from human-human interaction (x_v^*), and will generally be larger (so long as victim and injurer precautions are weak structural substitutes).

Finally, the first best social cost in the mixed autonomous/human world *need not be lower than in the absence of AV technology.* To see this, note that note that the expected social cost of victim-injurer interactions for the α-fraction of the injurer population without AV is $\Phi(x_i^*, x_v^*)$, whereas expected social cost with AV is $c(\hat{\gamma}_i) + (1 - \hat{\gamma}_i) \cdot \Phi(0, \hat{x}_v | \eta, \gamma_i)$. Depending on the characteristics of $c(\gamma_i)$, it is possible that the socially optimal domain for AV technology $\hat{\gamma}_i$ is relatively small, and thus a significant fraction injurer-victim interactions involving AV require that victims take all precautions unilaterally. It is easy to confirm that a sufficient condition for AV to be *suboptimal* relative to human-only interactions is:

$$\hat{\gamma}_i \leq \overline{\gamma} \equiv 1 - \frac{\Phi(x_i^*, x_v^*)}{\Phi(0, \hat{x}_v | \eta, \gamma_i)} \tag{21}$$

Equivalently, a *necessary* condition for AV to be socially optimal relative to human-only interactions is if $\hat{\gamma}_i > \overline{\gamma}$. In other words, unless AV technologies are capable of confronting sufficiently many pathways, encouraging adoption of AV is socially undesirable.

3.2.2 Efficient Legal Rules

Finally, consider now whether it is possible to implement the socially optimal allocation of quality investments and precautions.

No Fault. Consider first a no-fault (NF) regime as applied to the AV-injuer / human victim interaction. Notably, such a regime is not (at least on first blush) as socially undesirable as its counterpart in human-human interactions, since NF causes the victim to bear all of the risk associated with harm. Once γ_i is determined, this allocation is in fact efficient since the AV's incentives to take precautions is no longer a constraint on the problem (only the domain of its capacities is). Consequently, the victim will internalize the social planner's problem as above: when the harm pathway is within the AV's domain ($\eta \leq \gamma_i$), the victim will (optimally) invest nothing in precautions; and when the harm pathway is outside the AV's domain ($\eta > \gamma_i$), the victim faces the same problem as the social planner in (16) above. At the same time, NF provides no incentives whatsoever for the injurer to invest in quality. Indeed, the injurer gains nothing by investing in quality (since all liability risk is borne by the victim), and the injurer optimally selects $\gamma_i^{NF} = 0$, incurring cost $c(0) = 0$, yielding total social cost of $\Phi(0, \hat{x}_v | \eta, \gamma_i)$ — the same as the NF regime with human-human interactions.

Strict Liability With a SL rule, just as before, the victim is perfectly insured against a harm, and he thus has no incentive to invest in precautions, setting $x_v = 0$. The AV-injurer, anticipating this behavior, takes it into account in installing quality, setting γ_i to solve:

$$\min_{\gamma_i} c(\gamma_i) + (1 - \gamma_i) \cdot p(0,0) \cdot H \tag{22}$$

The injurer's optimal level of quality under strict liability, or $\widetilde{\gamma}_i$, is characterized by:

$$c'(\widetilde{\gamma}_i) = p(0,0) \cdot H \tag{23}$$

Note that because $p(0,0) > p(0, \hat{x}_v)$, the injurer installs an excessive expansive pathway domain for the AV technology–a direct by-product of the absence of incentives by the victim to take precautions.

Negligence Family Finally, consider the family of negligence rules, which (as noted above) had numerous candidates capable of implementing first best. An observation that immediately emerges from this setup is that it is no longer coherent to tie the injurer's negligence to her precautions while driving, since that function is now algorithmic, and the AV has either infinite capacity to avoid accidents for pathways in its domain or none at all). However, a first-best injurer-based negligence standard is coherent at the design phase – effectively a form of products liability. Consider, for example, a fault-based products liability regime that imposes a fault threshold of γ_i^{PL}, the satisfaction of which by the injurer puts all risks on the victim (the functional equivalent of no-fault for AV technologies meeting the products liability threshold). A natural question to ask is whether first-best is implementable under this type of real when the fault standard is pegged against socially optimal quality, so that $\gamma_i^{PL} = \hat{\gamma}_i$. As it turns out, the answer to this question is a "maybe" at best. To see this, consider first the victim's objective when the injurer meets the prescribed first-best negligence standard setting $\gamma_i = \hat{\gamma}_i$, and the harm pathway is within the AV's domain ($\eta \leq \gamma_i$). Here the victim will optimally expend no resources on precautions, since the AV technology eliminates the risk of an accident completely. In contrast, suppose the harm pathway is outside of the AV's domain (so that

$\eta > \gamma_i$). Here, a above, the victim solves a problem identical to (16), and he will (optimally) expend $\hat{x}_v$ on precautions.

As for the injurer, things are slightly more complicated: Anticipating the victim's behavior in pathways outside the AV's domain ($x_v = \hat{x}_v$), the injurer sets γ_i to minimize costs, solving:

$$\min_{\gamma_i} \left[\Omega_i\left(\gamma_i | PL, \eta, \gamma_i\right)\right] \Leftrightarrow \min_{\gamma_i} \left[c\left(\gamma_i\right) + \left\{ \begin{array}{ccc} \left(1-\gamma_i\right)\cdot p\left(0,\hat{x}_v\right)\cdot H & if & \gamma_i < \hat{\gamma}_i \\ 0 & & else \end{array} \right.\right] \tag{24}$$

Note that when $\gamma_i \geq \hat{\gamma}_i$, the seller's cost is strictly increasing in γ_i, and thus she would never expend more than $\hat{\gamma}_i$ in this region. Moreover, the downward discontinuity in the seller's cost at $\gamma_i = \hat{\gamma}_i$ implies that $\hat{\gamma}_i$ is a local minimum. In contrast, when $\gamma_i < \hat{\gamma}_i$, the seller's marginal cost of installing quality is:

$$\Omega_i'\left(\gamma_i | PL, \eta, \gamma_i\right) = c'\left(\gamma_i\right) - p\left(0, \hat{x}_v\right) H \tag{25}$$

By contrast, recall that the marginal social cost of quality investments is strictly lower:

$$\Omega'\left(\gamma_i | \hat{x}_v, , \eta, \gamma_i\right) = c'\left(\gamma_i\right) - p\left(0, \hat{x}_v\right) H - w_v \hat{x}_v < \Omega_i'\left(\gamma_i | PL, \eta, \gamma_i\right), \tag{26}$$

and consequently the seller's optimization problem must have another local minimum at some $\widehat{\hat{\gamma}}_i < \hat{\gamma}_i$, characterized by:

$$c'\left(\widehat{\hat{\gamma}}_i\right) = p\left(0, \hat{x}_v\right) H \tag{27}$$

Which of the two local minima ($\widehat{\hat{\gamma}}_i$ or $\hat{\gamma}_i$) is a global minimum depends on the functional forms of $c(.)$ and $p(.)$. In situations where $\hat{\gamma}_i$ is relatively "high" (because, say, victim precautions are expensive[33]), the benefit to the injurer associated with the payoff discontinuity in meeting the fault standard grows modest, and it may become optimal for the injurer simply to miss the fault standard deliberately by installing $\widehat{\hat{\gamma}}_i < \hat{\gamma}_i$.

To the extent that a negligence standard fails to incentivize efficient investments in AV quality, however, it is possible to adjust damages to compensate. In particular, it is easily confirmed that a products-liability negligence rule can ensure optimal behavior by the injurer if it "inflates" damages by

[33] Note that the injurer accounts for only part of the benefit of investment in quality — the reduced exposure to damages. She does not account for the reduced cost of victim precautions on the margin as she invests in product quality.

an additional term $\frac{w_v \hat{x}_v}{p(0,\hat{x}_v)}$, so that a negligent AV-injurer also must bear the full expected value of the victim's precautions, and the total award in a successful products liability suit is $\left(H + \frac{w_v \hat{x}_v}{p(0,\hat{x}_v)}\right)$. With this alteration (and assuming the fault standard remains at the first best of $\hat{\gamma}_i$), the injurer's cost minimization problem becomes solving:

$$\min_{\gamma_i} \left[c(\gamma_i) + \left\{ \begin{array}{cc} (1-\gamma_i) \cdot p(0,\hat{x}_v) \cdot \left(H + \frac{w_v \hat{x}_v}{p(0,\hat{x}_v)}\right) & if \;\; \gamma_i < \hat{\gamma}_i \\ 0 & else \end{array} \right. \right] \tag{28}$$

As before, regardless of the injurer's investment, whenever $\eta > \gamma_i$ the victim will implement precautions $x_v = \hat{x}_v$. Anticipating this behavior, the injurer's marginal cost of quality investment when $\gamma_i < \hat{\gamma}_i$ becomes:

$$\begin{aligned} \Omega_i'(\gamma_i|\hat{x}_v) &= c'(\gamma_i) - p(0,\hat{x}_v) H - w_v \hat{x}_v \\ &= \Omega'(\gamma_i|\hat{x}_v), \end{aligned} \tag{29}$$

thereby aligning the injurer's ex ante cost with social ex ante cost, and ensuring that the injurer optimally installs quality of $\hat{\gamma}_i$. (Note that because the injurer finds it optimal to satisfy the negligence standard, the victim is not in a position to collect the supra-compensatory award and thus his incentives to expend $\hat{x}_i$ remain unaffected.

Alternatively, consider the "dual" to the above negligence rule – where the injurer incurs strict liability, but subject to contributory negligence defense as to the victim's precautions. Recall that under this regime, the victim has an automatic right to damages upon harm occurring, but the injurer is absolved if the victim's own precautions fall short of the contributory negligence threshold x_v^{CN}. Similar to the above, it turns out that if fault threshold is set optimally at $x_v^{CN} = \hat{x}_v$ when $\eta > \gamma_i$, and if damages are inflated by $\frac{w_v \hat{x}_v}{p(0,\hat{x}_v)}$, then first best is once again implementable in equilibrium. To see this, suppose first that the injurer expects the victim to engage in precautions of $\hat{x}_v$, thereby fobbing the entire liability risk on the injurer. The injurer's cost function becomes identical to that in (28), and thus it is optimal for the injurer to install quality level $\hat{\gamma}_i$. Anticipating this behavior, and assuming a pathway obtains that is outside of the AV's domain, the victim's cost minimization problem is:

$$\min_{x_v} \left\{ \begin{array}{cc} p(0,x_v) \cdot H + w_v x_v & if \;\; x_v < \hat{x}_v \\ w_v x_v - w_v \hat{x}_v & else \end{array} \right\} \tag{30}$$

Here it is clear that the victim can ensure that he bears total cost of zero by setting $x_v = \hat{x}_v$, while costs are strictly positive for all other values of x_v.

As was the case in the prior section, the analysis of the negligence family yields the same result for other variations, including (a) a "nested" rule imposing negligence on product design but subject to contributory negligence, or (b) either a negligence or strict liability rule subject to comparative negligence (pure or modified).

3.3 Synthesis

Table 2 below summarizes the results from the model thus far.

	Human-Only Baseline	Human-AV Mixed System (Symmetric Information) AV Injurers (α)	Human Injurers (1-α)
No Fault (NF)	Suboptimal	Suboptimal	Suboptimal
Strict Liability (SL)	Suboptimal	Suboptimal	Suboptimal
Ordinary Negligence as to Precautions	Optimal if fault standard set at injurer's first-best precaution level	Suboptimal (N/A)	Optimal if fault standard set at injurer's first-best precaution level
Ordinary Negligence as to Precautions s.t. Contrib. / Compar. Negligence	Optimal if fault standard set at injurer's first-best precaution level	Suboptimal (N/A)	Optimal if fault standard set at injurer's first-best precaution level
Ordinary Negligence as to Product Design	Suboptimal (N/A)	Optimal if fault standard set at injurer-designer's first-best product design and punitive damages allowed	Suboptimal (N/A)
SL s.t. Contrib. / Compar. Negligence as to Precautions	Optimal if contributory fault standard set at victim's first-best precaution level	Optimal if contributory fault standard set at victim's first-best precaution level and punitive damages allowed	Optimal if contributory fault standard set at victim's first-best precaution level
Ordinary Negligence as to product design s.t. Contrib. / Compar. Negligence as to Precautions	Suboptimal	Optimal if fault standard set at injurer-designer's first-best product design and punitive damages allowed	Suboptimal

Table 2. Synthesis of Results Comparing Human-Only to Human/AV Interactions

The rows of the table consider focal forms of liability, while the columns of the table consider various permutations of interaction, involving human only versus human/AV mixed interactions. In the latter cases, victims are assumed to have complete information about injurers' adoption of AV, its capabilities and the realized pathway / state of the world. Note that neither strict liability nor no-fault regimes fare any better in an AV-Human environment. Moreover, ordinary negligence (as to precautions) is also not up to the task in a hybrid environment, while an ordinary negligence rule as to product design likely would be. The only framework that tends to work across domains of those in Table 1 is SL subject to a contributory negligence defense. Even here, however, the attributes of the fault standard for victims would be different when comparing human-human interactions and human-AV interactions. In many plausible settings (such as when precautions are partial substitutes for one another), the victim's prescribed fault standard is *more* demanding in the human-AV context than it would be in a human-human context.

4 Extensions and Robustness

This section [still in draft] considers a variety of extensions to the model, altering some of its more restrictive assumptions.

4.1 Alternative Information Environments

In the previous section, victims were assumed to be able to (a) discern AV from human driven vehicles; (b) determine the realized pathway η before acting, and (c) observe the degree of safety installed on an AV. Such assumptions allowed one to subdivide both social optima and legal rights and obligations, making them fully contingent on both whether they involve autonomous or human injurers as well as the degree of quality installed on AVs. In more realistic setting, victims may not have the capacity to observe these factors in real time – even if they are potentially verifiable after the fact. This section, therefore, briefly considers how the results from the prior section would change if certain of the above factors were not observable to victim when he must act. When the victim cannot observe characteristics about the injurer, he must instead make conjectures about such characteristics, responding optimally given those conjectures. An equilibrium thus minimally requires that those conjectures be borne out along the equilibrium path.

Consider first the case where the victim is able to discern AVs from human injurers and the realization of η, but he is unable to discern the quality level installed by an autonomous injurer, or γ_i. As it turns out, relaxing this informational assumption has minimal effect. In the NF regime, for example, it is common knowledge that the injurer has no incentives to provide quality, and the victim (accurately) conjectures such. Similarly, in the SL regime, the victim is fully insured regardless of what value of η obtains. Consequently, her behavior is the same (set $x_v = 0$) regardless of her conjecture. Under a simple negligence rule (with enhanced damages, as above), the plaintiff conjectures (accurately) that the defendant finds it optimal to satisfy the fault standard by installing $\hat{\gamma}_i$, and thus the plaintiff bears all the liability risk whenever the realized η exceeds the conjectured quality of the AV technology and installs $x_v = \hat{x}_v$. Finally, in the SL regime with a contributory negligence defense (also with enhanced damages), the victim optimally meets the negligence threshold whenever realized η exceeds the conjectured quality, setting $x_v = \hat{x}_v$. This conjecture is once again borne out in equilibrium. Once again, both ordinary negligence and SL with contributory negligence (among others) are able to implement efficient precautions and quality provision.

Things grow somewhat more complicated when the victim cannot discern the realization of η. (Or equivalently, the victim is able to determine the nature of the harm pathway, but does not understand how different values of η correspond to the AV's conjectured quality.) Here, when the victim observes (or accurately conjectures) an AV quality level γ_i, he is at best able to deduce that $\eta \leq \gamma_i$ with probability γ_i and $\eta > \gamma_i$ with complementary probability $(1 - \gamma_i)$. When the victim is constrained in this way, it is no longer feasible to target victim-side precautions only to those contexts where the pathway exceeds the AV's capability – rather, the victim must invest in a "one-size-fits-all" precaution, which will (with probability γ_i) be duplicative of and redundant to the AV's technology. Imposing this constraint on the interim social costs yields:

$$\begin{aligned} \Phi\left(x_v|\gamma_i\right) &= \gamma_i \cdot \left(w_v \cdot x_v\right) + \left(1-\gamma_i\right) \cdot \left(p\left(0, x_v\right) \cdot H - w_v \cdot x_v\right) \qquad (31) \\ &= \left(1-\gamma_i\right) \cdot \hat{p}\left(x_v\right) \cdot H + w_v \cdot x_v \end{aligned}$$

And the optimal victim precaution is now equal to $\tilde{x}_v$, characterized by the condition:

$$\left(1-\gamma_i\right) \frac{\partial \hat{p}}{\partial x_v} \cdot H + w_v = 0 \qquad (32)$$

It is easily confirmed that whenever $\gamma_i > 0$ we must have $\tilde{x}_v < \hat{x}_v$, and thus it is socially optimal to reduce victim-side precautions due to their wasteful duplication of the algorithm's effort. Backing up, the socially optimal level of quality provision is $\tilde{\gamma}_i$, where:

$$\Omega\left(\gamma_i|\tilde{x}_v\right) = c\left(\gamma_i\right) + \Phi\left(0, \tilde{x}_v|\gamma_i\right) \qquad (33)$$

The first order conditions associated with cost minimization

$$c'(\gamma_i) = \hat{p}(x_v) \cdot H \tag{34}$$

which yields a unique optimal quality level $\tilde{\gamma}_i$. It is straightforward to show[34] that $\tilde{\gamma}_i > \hat{\gamma}_i$, and thus when the victim cannot discern η, the first-best allocation increase AV quality even further than in the baseline case. Consequently, when the victim cannot discern η, the social planner requires even greater capacity out of the AV technology than would otherwise be the case. Beyond this point, however, the negligence family of rules (with escalated damages) is able to implement the social optimum.

Finally, suppose that victims were unable to discern even whether the injurer is a human or AV. (As noted in the introduction, this is an increasingly likely scenario). In this case, the victim must once again make an "untailored" investment in precaution, which now cannot be conditioned on either η or the type of injurer. In this case, ex ante social costs become:

$$\begin{aligned} \Omega(x_i, x_v, \gamma_i) &= (1-\alpha) \cdot [p(x_i, x_v) \cdot H + w_i \cdot x_i] \\ &\quad + \alpha \cdot [c(\gamma_i) + (1-\gamma_i) \cdot \hat{p}(x_v) \cdot H] \\ &\quad + w_v \cdot x_v \end{aligned} \tag{35}$$

The first order conditions for the victim, AV injurer, and human injurer behavior are:

$$\begin{aligned} H \cdot \frac{\partial p}{\partial x_i} + w_i &= 0; \\ H \cdot \left(\frac{\partial p}{\partial x_v}(1-\alpha) + \frac{d\hat{p}}{dx_v}\alpha(1-\gamma_i) \right) + w_v &= 0; \\ c'(\gamma_i) - \hat{p}(x_v) \cdot H &= 0 \end{aligned} \tag{36}$$

Note that the optimality condition for the human injurer's precautions is the same as in the baseline model. However, the victim's is now a probabilistic amalgam across different scenarios. Except for special cases, the optimal precaution levels for both the victim and the human injurer will be different than in the baseline model. This is significant, because it suggests that the addition of AV technology can affect social optimality goals even in contexts where no actor is using AV technology.

[34] To see this, simply compare the first order condition in the text to the one in the baseline model (where $c'(\gamma_i) = \Phi(0, \hat{x}_v | \eta, \gamma_i)$, and note that $\hat{p}(x_v) \cdot H < \Phi(0, \hat{x}_v | \eta, \gamma_i) \equiv$ $, \hat{p}(x_v) \cdot H + w_v \hat{x}_v$

4.2 Other Extensions

Although not yet included in this draft, there are several other possible extensions of this framework that might be pursued. They include the following:

- AVs with Human Monitors (as in Uber/Tempe case)
- Activity Levels
- Endogenous Adoption of AV Technologies
- Heterogeneous Types
- Risk Aversion
- Bilateral Harms
- AV manufacturer/sellers distinct from purchaser/owners

5 Conclusion

[TBA - Still in Draft.]

6 References

1. American Bar Association, (2009). "Comparative/Contributory Negligence and Joint & Several Liability: A State-by-State Summary."

2. Anderson, James M., Nidhi Kalra, Karlyn D. Stanley, Paul Sorensen, Constantine Samaras and Oluwatobi A. Oluwatola (2014). Liability Implications of Autonomous Vehicle Technology, in Autonomous Vehicle Technology: A Guide for Policymakers (Published by the RAND Corporation).

3. Brown, J.P. (1973). "Toward an economic theory of liability". Journal of Legal Studies 2, 323–349.

4. Calabresi, G. (1970). The Costs of Accidents (Yale University Press).

5. Cooter, Robert & Thomas Ulen (2012). Law and Economics (6th Ed.).

6. Craswell, R., Calfee, J.E. (1986). “Deterrence and uncertain legal standards”. Journal of Law, Economics, & Organization 2, 279–303.

7. Engstrom, Nora (2018). “When Cars Crash: The Automobile’s Tort Law Legacy.” 53 Wake Forest L. Rev. 293.

8. Geistfeld, Mark A. (2017). “A Roadmap for Autonomous Vehicles: State Tort Liability, Automobile Insurance, and Federal Safety Regulation” 105 Cal. L. Rev. 1611.

9. Holmstrom, Bengt (1982). “Moral Hazard in Teams.” The Bell Journal of Economics Vol. 13, No. 2 pp. 324-340.

10. Hylton, Keith N. (2013). The Law and Economics of Products Liability, 88 Notre Dame L. Rev. 2457.

11. Keeton, W. Page (1969). Manufacturer’s Liability: The Meaning of "Defect in the Manufacture and Design of Products, 20 SYRACUSE L. REV. 559.

12. Kroll, Elliot & James Westerlind (2012). Survey of Damage Laws of the 50 States including the District of Columbia and Puerto Rico (Arent Fox LLP).

13. Maryland Department of Legislative Services (2004). “Negligence Systems: Contributory Negligence, Comparative Fault, and Joint and Several Liability.”

14. Mashaw, Jerry & David Harfst (1990). “The Struggle for Auto Safety.”

15. Rabin, Robert L. (2005). “Past as Prelude: The Legacy of Five Landmarks of Twentieth-Century Injury Law for the Future of Torts,” in EXPLORING TORT LAW (M. Stuart Madden ed.).

16. Shavell, S. (2007). “Liability for Accidents” in Handbook of Law and Economics, Volume 1 (A. Mitchell Polinsky & Steven Shavell eds.)

17. Shavell, S. (1987). Economic Analysis of Accident Law (Harvard University Press).

18. Shavell, S. (2004). Foundations of Economics Analysis of Law (Harvard Press).

19. Talley, E. (2018). "Is the Future of Law a Driverless Car? Assessing How the Data Analytics Revolution Will Transform Legal Practice." Journal of Inst. & Th. Econ., pp. 183-205.

20. Turner, Fox, and Debra Richardson (2002). "Strict Products Liability for Software Code Defects." Proceedings of the Third International Conference on Law and Technology (ISLAT).

21. Villasenor, John (2014). "Products Liability and Driverless Cars: Issues and Guiding Principles for Legislation." (Brookings report).