Liability for Robots

August 15, 2021

Abstract

In robot torts, robots carry out activities that are partially controlled by a human operator. The care of the operator and the prospective victim can reduce the likelihood of an accident. The problem for a policymaker is to design liability regimes that incentivize both the operator's and victim's optimal care, while inducing manufacturers to improve the safety of robots by internalizing the expected cost of non-negligent accidents. In this paper, we suggest the possibility of blending negligence-based rules for robot operators and their potential victims, and strict liability rules robots by manufacturers, to create care incentives and R&D incentives for developing safer robots. We refer to these regimes as rules of "manufacturer residual liability." By making both operators and victims liable for accidents due to their negligence, the rule incentivizes them to act diligently. Moreover, by making the manufacturers residually liable for non-negligent accidents, the rule incentivizes optimal investments in R&D for robots' safety. In turn, this will minimize the market price of safer robots, driving unsafe technology out of the market. Thanks to the percolation effect of residual liability, operators (and victims) will also be incentivized to adopt optimal activity levels in robots' usage.

Keywords: liability; negligence; manufacturer residual liability; automated technology
JEL Codes: K13, K32

1 Introduction

The economic analysis of tort law assumes the existence of at least two human actors: an injurer and a victim (e.g., Shavell, 1980, 1987; Miceli, 1997, 2017). Nonetheless, this assumption becomes increasingly tenuous with the advancement of automated technologies (e.g., Shavell, 2020). Rather than still being the mere instruments of human decision-makers, machines *are* the decision-makers. Machine decision-making reduces the impact of human negligence but does not entirely eliminate the risk of accidents, creating fresh dangers in the form of machine errors and design limitations. Let us refer to these decision-making machines generally as "robots." The problem unique to decision-making machines is that the effects of their operation are not mediated by a human operator. For example, whereas a human driver can take precautions or undertake evasive maneuvers in the case of a blown tire, there is little that an average automobile consumer can do to prevent or mitigate the harm of defective code in the algorithm of a self-driving car. Thus, we are faced with the problem of creating laws to govern self-operating machines.

Since robots are insensitive to threats of legal liability, the question arises: how are we to regulate this new class of potential tortfeasors? This paper adds to the recent literature on the regulation of robots (Bertolini, 2013; Lemley and Casey, 2019; Casey and Lemley, 2019; Talley, 2019; Bertolini, 2020; Bertolini and Riccaboni, 2020; Kovač, 2020; Shavell, 2020; Epstein, 2021), stressing the need for a systematic understanding of how legal incentives—so carefully calibrated to induce *human* actors to exercise precautionary care—need to be reshaped for robots, but ultimately suggesting that no revolutionary legal solution is needed to tackle the problem of robot torts.

The need for such a theory to better understand robot torts is urgent, given that robots are already capable of driving automobiles and trains, delivering packages, piloting aircraft, trading stocks, and performing surgery with minimal human input or supervision. Engineers and futurists predict more revolutionary changes are still to come. How the law grapples with these emerging technologies will affect their rates of adoption and future investments in research and development. In the extremum case, the choice of liability regime could even extinguish technological advancement altogether. How the law responds to robot torts is thus an issue of utmost importance.

At the level of utmost generality, it is important to bear in mind that human negligence and machine error do not represent equivalent risks. The social cost of machine error promises to be drastically lower than that of human negligence. We should therefore welcome the development of robot technology. Even if there was nothing that the law could do to reduce the risk of robot accidents, merely encouraging the transition to robot technology alone would likely effect a dramatic reduction in accident costs.

Our key analytical move is to cluster robot-generated accidents within the realm of fault-based liability rather than products liability. This analytical move follows straightforwardly from the observation that robots are no longer tools operated by humans, but rather they act as independent agents performing activities in the place of humans. Contrary to ordinary tools used by a human operator, robots serve as a replacement to the decision-making by a reasonable person.[1] By extending Shavell's (1987) standard economic model of accident law, we suggest and analyze the possibility of blending negligence-based rules and strict liability rules to create care incentives for robot operators and their potential victims, and R&D incentives for the development of safer robots by manufacturers.

A splicing of existing legal rules into a regime that we call "manufacturer residual liability" may help to pursue these objectives. Under a manufacturer residual liability regime, primary liability falls upon either the robot operator or the victim, and residual liability—the assignment of the accident cost when *neither* of those parties is negligent—falls upon the manufacturer.[2] In the context of robot torts, the assignment of residual liability to the third-party manufacturer offers several advantages over simple negligence and strict liability, and it accomplishes four objectives of a robot tort regime: (1) incentivize efficient human precautionary care, (2) incentivize efficient activity levels, (3) incentivize investment in the research and development of safer robots, and (4) incentivize the adoption of the newer, safer technology.

This paper comprises five sections. Section 2 reviews the relevant prior scholarly literature on the need to rethink remedies for robot torts and current law on robot torts in different industries and applications. Section 3 provides an exposition of our rule of manufacture residual liability. A formal model is used to study the incentives created by this rule. Section 4 unpacks the implications of our

[1] For example, an individual can "tell" a Google self-driving car to take him/her home, but has only limited control on how the car will accomplish that task. Unsurprisingly, in the context of self-driving cars, the term *driver* is meant to include a *corporation-driver* (Smith, 2014). States such as Nebraska have already adopted a broad definition of the term driver, which is "to *operate* or be in the actual physical control of a motor vehicle" (Neb. Rev. Stat. §60-468.), where *operate* has been defined by courts as including any mechanical or electrical agency that sets the vehicle in action or navigates the vehicle. Similarly, the National Highway Traffic Safety Administration has stated that the self-driving system can be considered the driver of the vehicle (National Highway Traffic Safety Administration, 2013).

[2] Our concept of "manufacturer residual liability" is distinct from the products liability concept of "residual-manufacturer liability" introduced by Hay and Spier (2005). In Hay and Spier, manufacturers become liable in case of the *insolvency* of product users who caused harm to a victim (e.g., shortfall in damages for the use of dangerous products such as firearms). Their concept of "residual-manufacturer liability" applies to the general class of strict liability rules, under which the liability *does not* depend upon the precautions taken by manufacturers and/or consumers. Instead, our "manufacturer residual liability" applies to the class of negligence rules, under which liability *does* depend upon the precautions taken by operators and victims (see related discussions in Sections 2 and 3.1). The manufacturer becomes liable when an accident occurs and both the operator and victim exercised due care. Further, again in contrast to Hay and Spier (2005), the efficiency of our rule (in terms of optimal care incentives and safety investments) does not depend upon the operator's financial assets. It rather holds for the broader class of harms caused by non-negligent operators to others. Given the different premises and scopes of the two rules, the incentives that they generate are also clearly different, as discussed throughout the paper.

model. Finally, Section 5 concludes with some remarks on practical implementation and suggestions for future research.

2 Background

The law of robot torts is currently at a nascent stage of development. There exists no general formulation of liability in case of accidents caused by robots, although some legislatures have attempted to anticipate some of the issues that could arise from robot torts. Section 2.1 discusses the novel problems posed by robot accidents, and the reasons why robots rather than other machines need special legal treatment. Section 2.2 reviews the prior literature, suggesting how these problems could be addressed. Section 2.3 presents the current legal approaches to dealing with robot accidents.

2.1 Rethinking Legal Remedies for Robots

In an early article in *Science*, Duda and Shortliffe (1983) argued that the difference between a computerized instrument and a robot is intent. A computerized instrument—such as a computer program—is intended to aid human choice, while a robot becomes an autonomous knowledge-based, learning system, whose operation rivals, replaces, and outperforms that of human experts (Duda and Shortliffe, 1983, pp. 261–268). Thanks to the dynamic nature of the decision algorithm that drives their behavior, robots take into account the new information gathered in the course of their operation and dynamically adjust their way of operating, learning from their own past actions and mistakes (Giuffrida et al., 2017; Giuffrida, 2019).

In the face of the superior decision-making skills of a robot, the relationship between a robot and its operator is different from the relationship between an ordinary tool and its user. As the skills of a robot increase, the need and desirability of human intervention decreases.[3] Even though there may be special circumstances in which human judgment may outperform robots, robots outperform

[3] Two new modes of programming that differ from the traditional algorithmic programming of robots—"machine learning" and "genetic and evolutionary programing"—have further expanded the horizons in the evolution of artificial intelligence. With these programming modes, robots operate with a range of programs that randomly compete against each other, and only the variations of the program that carry out tasks best will survive, while others will die (a "survival of the fittest" programming approach). The surviving programs will replicate themselves, making slight modifications of their "genes" for the next round of tasks (Michalski, 2018). See also Michalski's (2018) discussion about companies that invested in robots capable of building other improved robots, thus putting human creators one more step away from future prospective victims.

humans in most situations. Humans defer to the superior skills of a robot and delegate important decisions to them (Casey, 2019). However, as robots increase their skills, their "thinking" becomes more "inscrutable," falling beyond the human computational capacity (Michalski, 2018).[4] Given the opacity of the robot's decisions, it is very difficult—and often unwise—for operators to second-guess and override the decisions of a robot (Lemley and Casey, 2019).

The high complexity of the decision algorithm and the dynamic adjustment of the programming in unforeseen circumstances are what make robots different from other machines and what—according to many scholars—call for special legal treatment and a new approach to modeling accidents.[5] Several legal and economic scholars across the world have argued for the need to rethink legal remedies as we apply them to robot torts (e.g., Matsuzaki and Lindemann, 2016; Lemley and Casey, 2019; Talley, 2019; Shavell, 2020).[6] The proposed legal solutions to robot torts differ across jurisdictions (e.g., Europe *versus* Japan; Matsuzaki and Lindemann, 2016),[7] yet the common awareness is that as the level of robot autonomy grows, under conventional torts or products liability law it will become increasingly difficult to attribute responsibility for robot accidents to a specific party. This problem is what Matthias (2004) calls the "responsibility gap."[8] Matsuzaki and Lindemann (2016) note that in both the EU and Japan, the belief is that product liability's focus on safety would impair the autonomous functioning of the robot and slow down the necessary experimentation with new programming techniques. In a similar vein, in their US-focused article titled "*Remedies for Robots*," Lemley and Casey wrote: "Robots will require us to rethink many of our current doctrines. They also offer important insights into the law of remedies we already apply to people and corporations" (Lemley and Casey, 2019,

[4]Mulligan (2017) refers to these as "black box algorithms," ones that not even the original designers and programmers can decipher. "Machine learning" and "genetic and evolutionary programing" (see supra note 3) have further increased the complexity and opacity of the robot's decision-making process.

[5]We thank Professor Geoffrey M. Hodgson for encouraging us to elaborate on the discussion that follows.

[6]For a survey of the difficulties that legal scholars face when attempting to apply existing legal rules to robot torts, see Chopra and White (2011, Chapter 4; pp. 119–152).

[7]In their comparative study, Matsuzaki and Lindemann (2016) show that both the legal framing and the concrete solutions to robot torts differ between Europe and Japan, especially in the legal construct of the robot as an "agent" of the operator. While the European regulation debate explicitly addresses the degree of machine autonomy and its impact on legal institutions, this is not the case in Japan. See also, e.g., Leis (2006); MacDorman et al. (2009); Šabanović (2014).

[8]Specifically, Matthias (2004) wrote: "The rules by which [robots] act are not fixed during the production process, but can be changed during the operation of the machine, by the machine itself. This is what we call machine learning. [...] [T]he traditional ways of responsibility ascription are not compatible with our sense of justice and the moral framework of society because nobody has enough control over the machine's actions to be able to assume the responsibility for them. These cases constitute what we will call the responsibility gap" (Matthias, 2004, p.177).

p. 1311). Robots amount to a paradigmatic shift in the concept of instrumental products, which—according to Talley (2019) and Shavell (2020)—renders products liability law unable to create optimal incentives for the use, production, and adoption of safer robots as it is currently designed.

Under the manufacturer residual liability rules proposed in this paper, a manufacturer's liability would arise for two separate sources of accidents caused by robots: "malfunctions" and "design limitations." Malfunctions occur when the robot is not able to execute the intended decision of the algorithm or when bugs in the algorithm cause unintended behavior. Liability for "malfunctions" could be dealt with by ordinary products liability law, allowing victims to sue manufacturers directly, or allowing operators to sue manufacturers in subrogation when operators face direct liability under conventional tort law. One of the challenges in the regulation of robots—which is the focus of our analysis—concerns accidents caused by "design limitations," i.e., accidents that occur when the robot encounters a new unforeseen circumstance that causes it to behave in an undesired manner. The environments in which robots operate are complex and the problems that may be unforeseen by robots' designers are innumerable: design perfection is the unachievable target of technological progress.

Although not every accident arising from the actions of a robot can be attributed to a manufacturing defect, hypothetically most accidents could be avoided with greater investments in R&D and/or safety updates. For example, the algorithm of a self-driving car could not "know" that a particular stretch of road is unusually slippery, or that a certain street is used by teenagers for drag racing on weekends. Under conventional products liability law, we could not hold a manufacturer liable for a design defect for not having included that specific information in the software. Failing to account for every special circumstance cannot be regarded as a design flaw. However, we could design rules that might keep incentives in place for manufacturers to narrow the range of design limitations through R&D investments. In our example, we may be able to incentivize manufacturers of self-driving cars to design cars that can "learn" information and share their dynamic knowledge with other cars to reduce the risk of accidents in those locations. As it will be shown in Section 3.2, one of the effects of our rules of manufacturer residual liability is to incentivize manufacturers to constantly improve the design of their robots, while keeping in place all of the other parties' primary incentives.

Calo (2015) argued that for the first time robots combine the ability to process information and the capacity to do physical harm. Rachum-Twaig (2020) added that due to this ability to accomplish both useful and harmful tasks, robots are increasingly perceived in society as social actors. Although legal scholars recognize that robots are mere physical instruments and not social actors, some have argued that from a pragmatic perspective, granting them a legal personality status—similar to corporations—might address some of the responsibility problems addressed above.[9]

The idea of attributing legal personhood to robots has been entertained in both Europe and the US. The European Parliament has proposed the creation of specific status for autonomous robots, a third type of personhood between natural personhood and legal personhood, called "electronic personhood" (European Parliament, 2017). The mechanics of how the electronic personhood of robots would operate is broadly presented by Bertolini and Episcopo (2021): "Attributing legal personhood to a given technology, demanding its registration and compliance with public disclosure duties, minimal capital and eventually insurance coverage would turn it into the entry point for all litigation, easing the claimants' position" (Bertolini and Episcopo, 2021, p.14). The idea of giving some form of legal personality to robots has also been voiced in the US (Jones, 2018; Kop, 2019; Carroll, 2021), although it has never advanced to the legislative level.

Many challenges would arise in the application of existing tort instruments to robots with electronic personality. Traditional legal rules refer to human-focused concepts such as willfulness, foreseeability, and the duty to act honestly and in good faith, concepts that no longer fit the new realities involving robots. Unlike humans, robots are insulated from self-interested incentives, which is intrinsically a good thing. However, the robots' insulation from self-interested incentives can at times be a double-edged sword. Robots are not deterred by threats of legal or financial liability, since their personal freedoms and wealth are not at stake. To cope with this shortcoming, scholars and policymakers have investigated the possibility to make robots bearers of rights and duties, and holders of assets like corporations (Giuffrida, 2019; Bertolini, 2020; Bertolini and Riccaboni, 2020).

In our paper, we bracket off the questions raised in the electronic personhood literature, and proceed to study liability for robots as they are currently framed under the law. In Section 5, we recom-

[9] As Carroll (2021) points out, due to the"black box" problem, nobody understands how a robot thinks, which creates a difficult accountability gap between manufacturers, operators, and victims. The attribution of legal personality to a robot thus becomes a possible way to fill the accountability gap: "the legal framework that the US should ultimately adopt for the liability of self-driving cars is the notion of electronic legal personhood" (Carroll, 2021, p.2).

mend future extensions of our results to consider how rules of manufacturer residual liability could be applied to robots with legal personality and independent assets. In the following, we also bracket off the many interesting philosophical questions that commonly arise in the contemplation of autonomous robots' decision-making. For instance, a self-driving car may be faced with a situation where the vehicle ahead of it abruptly brakes, and the robot must choose whether to collide with that vehicle or swerve onto the sidewalk, where it risks hitting pedestrians. Alternatively, a robot surgeon may be forced to make split-second decisions requiring contentious value judgments. For example, should the robot choose a course of action that would result in a high chance of death and low chance of healthy recovery, or one that would result in a lower chance of death but a higher chance of survival with an absysmally low quality of life? While these moral questions are serious and difficult, we exclude them from our inquiry because we do not consider them critical for the solution to the *incentive* problem that we are tackling.[10] As a practical matter, it cannot seriously be entertained that the design of rules governing such a critical area of technological progress should be put on hold until philosophers "solve" the trolley problem or the infinitude of thought experiments like it. Second, even if "right answers"' exist to the ethical problems that a robot may face, its failure to choose the "morally correct" course of action in some novel circumstance unanticipated by its designers can be construed by courts or lawmakers as a basis for *legal* liability. The objective of tort law is to minimize the social cost of accidents, and if the compliance with virtuous conduct in ethical boundary cases helps to accomplish that social objective, ethical standards should be incorporated into the legal standards of due care. Finally, if it is mandated as a matter of public policy that a certain approach to moral problems should be implemented, then this can be effected by direct regulation of robot manufacturing, outside of rules of tort liability.

2.2 Related Literature

In the economic analysis of tort law, there have been several forays into analyzing self-driving car torts (e.g., Fagnant and Kockelman, 2015; Roe, 2019; Shavell, 2020). Most robot-generated accidents have been analyzed under a framework of products liability, assigning strict liability to manufacturers

[10]For some discussion of ethical issues on the regulation of robots, see Giuffrida (2019).

(e.g., Ben-Shahar, 2016; Crane et al., 2017; Evas et al., 2018; Abraham and Rabin, 2019). However, for the reasons that we have already discussed, we believe that the conceptual problem is more general. Approaching the problem from a products liability perspective *assumes* many of the answers. As will be further shown in Section 3.2, our rule of "manufacturer residual liability" is not a rule of products liability. Manufacturers are strictly liable when operators and victims are not negligent, regardless of manufacturing or design defects, as it would be generally required under products liability.

Two existing papers have approached the tort problem at a greater level of generality: Lemley and Casey (2019) and Talley (2019). Talley (2019) argues that standard negligence-based rules, coupled with doctrinal reforms and a reconceptualization of fault standards are able to provide optimal care and safety incentives in the case of self-driving car accidents. We believe that our approach of "manufacturer residual liability" is simpler and more practicable to optimally align incentives for all three parties in robot torts.

In a recent paper, Shavell (2020) proposes a rule of strict liability for automated vehicles and robot manufacturers with damages payable to the state. This is equivalent to a decoupling rule (see Polinsky and Che, 1991). Our rule substantially differs from Shavell's, as under our rule robot manufacturers face residual strict liability, and damages are paid to faultless victims rather than the state. The advantage of our rule is two-fold. In the first place, a rule that leads to payment of damages to the actual victims of the accident rather than the state serves a desirable corrective justice and compensatory function. Second, the ability to obtain compensation for the loss suffered gives non-negligent victims full incentives to activate the enforcement of the liability rule by engaging in litigation. As is well known (e.g., Polinsky and Che, 1991), these incentives may be diluted under decoupling rules, because in practice the injured party has no incentive to sue, obliterating the tort system's capacity to effect incentives.

Although the specific question of liability for robots has only been considered by a few contributions in the literature, two general phenomena characterizing robot torts are well studied in the prior literature: durable precautions and safer technologies. A durable precaution requires an upfront investment to reduce the probability of accidents occurring, whose effectiveness does not diminish with increasing activity levels. Research on the incentive effects of durable precautions is extensive (see, e.g., Grady, 1987; Shavell, 2008; Grady, 2009; Nussim and Tabbach, 2009; Mot and Depoorter, 2011; Dari-Mattiacci and Franzoni, 2014). [11]

Tort scholars have also studied incentives to upgrade to newer, safer technologies.[12] For example, Dari-Mattiacci and Franzoni (2014) analyzed the relationship between negligence standards and technology adoption. An implication of their research is that in cases of products liability, manufacturers—who are more sensitive to incentives to introduce harm-reducing innovations—should bear liability in case of accidents.

Durable precautions and upgrading technologies are both implicated in robot torts. Robot technology is the quintessential durable precaution. In the extremum case, a fully-automated robot renders manual precautionary care meaningless. It is also a safer technology. In most cases, we can assume that robots will be prevented from replacing human decision-makers by regulatory legislation unless they are shown to be safer than human decision-makers. The elimination of human decision-making has the potential to drastically reduce the rate of accidents.[13]

It is worth focusing specifically on self-driving cars. As the most salient robot technology on the horizon, the topic has attracted some theoretical speculation. Some legal scholars have questioned whether current tort rules present a barrier to the development and adoption of self-driving cars. Three approaches to the emergence of self-driving cars have been proposed: (1) current products liability law is still effective and adaptable to self-driving vehicles (Garza, 2011);[14] (2) a systematic review of the traditional tort system is needed to enable the development and production of safety technologies

[11] Grady (1987) first defined precautions as "durable" if they are long-lasting and require a single isolated measure to be taken and "non-durable" if they are short-lived and require repeated investments in care with each unit of activity. This distinction has subsequently been applied in several models: Shavell (2008) argued for preserving incentives to invest in durable precautions from changes in legal rules, Nussim and Tabbach (2009) revised the traditional model of unilateral accidents by distinguishing between durable and nondurable precautions, Mot and Depoorter (2011) focused on memory costs and showed that accidents involving non-durable technologies occur more frequently than those involving durable technologies, and Dari-Mattiacci and Franzoni (2014) showed that the impact of new technologies on expected harm should guide the definition of due care standards. See also Gilo and Guttel (2009) and Cooter and Porat (2013).

[12] The proposition that there exists in tort a duty to maintain up-to-date technology is well established in the caselaw. *The T.J. Hooper* (60 F.2d 737 [1932]).

[13] The National Highway Traffic Safety Administration's *Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey* (2015) attributes 94% of car accidents to human choices.

[14] Garza (2011) argued that products liability law in its present state may not need to be amended for autonomous vehicles. He claimed that despite criticism, the expected increase in manufacturer liability will not be a "dire concern" since self-driving cars will make automobile travel safer, leading to a net decrease in the costs of liability, insurance, and litigation. Alternatively, Mele (2013) argued that traditional negligence rules, together with regulatory changes could overcome the current legal barriers.

in a timely fashion (Calo et al., 2016);[15] and (3) no prediction is possible until the new technologies are introduced in the market (Graham, 2012).[16]

2.3 Current Legal Status

Robots are presently used in a variety of different settings. In some areas, they are already commonplace, while in other the technologies remain in their early stages (Księżak and Wojtczak, 2020). In this section, we survey some representative implementations to observe how legal rules have responded to the presence of robot actors to date.

2.3.1 Corporate Robots

In 2014, a Hong Kong-based venture capital fund appointed a robot to its board of directors. The robot—named "Vital"—was chosen for its ability to identify market trends that were not immediately detectable by humans. The robot was given a vote on the board "as a member of the board with observer status", allowing it to operate autonomously when making investment decisions. Although to our knowledge Vital in Hong Kong is the only robot benefiting from a board seat, and the recognition of personality to a robot does not extend to other jurisdictions, the World Economic Forum released a 2015 report where nearly half of the 800 IT executives surveyed expected additional robots to be on corporate boards by 2025 (World Economic Forum, 2015). At present, Hong Kong and the UK already allow the delegation of directors' duties to "supervised" robots (Möslein, 2018).

The adoption of robots in corporate boardrooms will unavoidably raise legal questions on the liability arising from directors' use of robots and for losses to corporate investors and creditors caused by robots' errors (Zolfagharifard, 2014; Burridge, 2017; Fox et al., 2019). As of today, these questions remain without proper answers and the regulation of corporate robots has been left within the

[15] Calo (2011) analyzed the commercial prospects of robotics in the United States, suggesting tort immunity for manufacturers to encourage the aggressive development of these technologies. He further proposed supplementing manufacturer immunity with markets for user insurance. Marchant and Lindor (2012) analyzed some legal tools that may protect manufacturers from the expected increase in liability and discussed the possibility that manufacturers' immunity may reduce the incentives to make incremental improvements in the safety of autonomous systems. See also Colonna (2012); Schellekens (2015).

[16] Graham (2012) emphasized the uncertainty surrounding the application of tort law to emerging technologies. He argued that the precise content of the legal rules for autonomous cars will remain unclear until these vehicles appear on public highways.

discretionary shield of corporate charters.

2.3.2 Aircraft Autopilot

Aircraft autopilot systems are among the oldest class of robot technologies. The earliest robot flight system—a gyroscopic wing leveler—was implemented as far back as 1909 (Cooling and Herbers, 1983, p. 693). After a century of development, autopilot technology has progressed to nearly full automation. Aircraft autopilot systems are presently capable of taking off, navigating to a destination, and landing with minimal human input.

The longevity of the autopilot technology in aviation affords us a clear exemplar of how the law can respond to the emergence of robot technology. Early treatment of autopilot cases was mixed. The standard for liability was not negligence, but rather strict liability. However, the cases were not litigated as a species of products liability. Aircraft and autopilot systems manufacturers were therefore rarely found liable (see *Goldsmith v. Martin* (221 F. Supp. 91 [1962]); see also Eish and Hwang, 2015; Cooling and Herbers, 1983). Relatively early onwards, it was established that operators (i.e., the airlines) would be held liable when an accident was caused by an autopilot system (see *Nelson v. American Airlines* (263 Cal. App. 2d 742 [1968])). There were two main reasons why aircraft and autopilot manufacturers were generally successful in avoiding liability: first, they were punctilious in crafting enforceable disclaimers and safety warnings, which effectively shielded them from products liability claims; and second, manufacturers aggressively litigated any claims against them, rarely settled, and thereby established favorable precedents (Leveen, 1983).[17]

The legal outcome is largely unchanged today. It remains the airlines—not the manufacturers—that are liable for harms caused by autopilot systems. However, although the result has not changed, the legal justifications have evolved. Products liability law has undergone a radical transformation since the early autopilot accident cases, yet manufacturers continue to successfully avoid liability, for two reasons. First, in order for a products liability claim to succeed, the risk of harm must be reasonably foreseeable. Present-day aircraft manufacturing is heavily regulated, and an autopilot system

[17] The regulation of autopilots and other aviation equipment in the EU and Japan is equally nuanced. See, e.g., "Easy Access Rules for Airworthiness and Environmental Certification (Regulation (E.U.) No 748/2012)" for the EU, and the "General Policy for Approval of Types and Specifications of Appliances" for Japan (available at https://www.mlit.go.jp/common/001111795.pdf).

that satisfactorily meets Federal Aviation Administration requirements is unlikely to be susceptible to any "reasonably foreseeable" risk of harm. Direct regulation thus pre-empts tort liability. Second, even when an autopilot system is engaged, pilots have a duty to monitor and override it if operation becomes unsafe.[18] The logic is that the human operator is legally responsible for anything that a robot does, because the human ultimately chooses to engage (and not override) the machine.

2.3.3 Self-Driving Cars

Self-driving cars are the most salient future use of robot technology. For quite some time, prototypes have demonstrated the feasibility of the technology, and fully autonomous vehicles are now part of the daily reality, from private cars to commercial taxi transportation, delivery robots, and self-driving trucks.[19] In September 2016, the Department of Transportation published the *Federal Automated Vehicles Policy*, providing legislative guidance for states contemplating the regulation of self-driving cars (National Highway Traffic Safety Administration, 2016). A growing number of jurisdictions have enacted laws regulating the use of self-driving cars. At present, in the US, 29 states and the District of Columbia have legislation relating to autonomous vehicles.[20] However, legislative efforts thus far have principally focused on determining whether an autonomous vehicle may be operated on public roads.[21] Few jurisdictions have attempted to address the tort issues relating to self-driving cars. The *Federal Automated Vehicles Policy* suggests various factors that lawmakers should consider when formulating a liability rule (National Highway Traffic Safety Administration, 2016, pp. 45–46):

[18] 14 Code of Federal Regulations § 91.3 ("The pilot in command of an aircraft is directly responsible for, and is the final authority as to, the operation of that aircraft.").

[19] The Society of Automotive Engineers (SAE) defines six levels of autonomy, ranging from 0 (fully manual) to 5 (fully autonomous). Most automakers currently developing self-driving vehicles are seeking level 4 autonomy, which does not require human interaction in most circumstances. Usually these vehicles are limited to routes or areas that have previously been mapped. See SAE International. 2021. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. https://www.sae.org/standards/content/j3016_202104/ (last updated April 30, 2021).

[20] The National Conference of State Legislatures maintains a database of state autonomous vehicle legislation, which is regularly updated. The searchable database is available at https://www.ncsl.org/research/transportation/autonomous-vehicles-legislative-database.aspx (last accessed June 4, 2020).

[21] For example, several states require that a self-driving car (i) satisfies the vehicular requirements of human-operated cars, (ii) is capable of complying with traffic and safety laws, and (iii) in case of a system failure, is capable of entering a "minimal risk condition," which allows the car to achieve a reasonably safe state (for example, by pulling over to the shoulder of the road when feasible). See, e.g., Iowa Code §321.515 (2020); La. Stat. Ann. §32:400.3 (2019); N.C. Gen. Stat. §20-401(h) (2020); Tenn. Code Ann. §§55-30-103 (2019).

> States are responsible for determining liability rules for HAVs ["highly automated vehicles"]. States should consider how to allocate liability among HAV owners, operators, passengers, manufacturers, and others when a crash occurs. For example, if an HAV is determined to be at fault in a crash then who should be held liable? For insurance, States need to determine who (owner, operator, passenger, manufacturer, etc.) must carry motor vehicle insurance. Determination of who or what is the "driver" of an HAV in a given circumstance does not necessarily determine liability for crashes involving that HAV. For example, States may determine that in some circumstances liability for a crash involving a human driver of an HAV should be assigned to the manufacturer of the HAV.
>
> Rules and laws allocating tort liability could have a significant effect on both consumer acceptance of HAVs and their rate of deployment. Such rules also could have a substantial effect on the level and incidence of automobile liability insurance costs in jurisdictions in which HAVs operate.

The few jurisdictions to address the problem of tort liability merely push the problem back. For example, Tenn. Code Ann. §55-30-106(a) (2019) states that "[l]iability for accidents involving an [Automated Driving System]-operated vehicle shall be determined in accordance with product liability law, common law, or other applicable federal or state law." Other states have enacted similarly opaque boilerplate that fails to delineate the applicable liability rule and define the legal "driver" in the context of self-driving cars as being the robot itself.

In the EU, driverless vehicle policy proposals have evolved into a multinational policy initiative, under the United Nations. EU member states—and other countries, including Japan and South Korea—have agreed to common regulations for vehicles that can take over *some* driving functions (e.g., mandatory use of a black box; automated lane keeping systems).[22] Nonetheless, unlike in the US, those countries currently do not have specific regulations for fully-automated cars.[23] In the UK, policies on driverless vehicles are still evolving, and press releases from the UK Department of Transport refer to a regulatory process that has been underway since the summer of 2020 and will reach a

[22] See ECE/TRANS/WP.29/2020/81 and "Addendum 156 – UN Regulation No. 157" of 4 March 2021.

[23] See, "UN Regulation on Automated Lane Keeping Systems is milestone for safe introduction of automated vehicles in traffic." Published June 24, 2020. Available at https://unece.org/transport/press/un-regulation-automated-lane-keeping-systems-milestone-safe-introduction-automated. Last accessed: July 2021.

point of greater completion in 2021 and the following years.[24]

In Japan, the Road Transport Vehicle Act and the Road Traffic Act were revised to account for the possibility of autonomous vehicles driving on public roads (Imai, 2019). Those revisions have significantly reduced the legal obstacles to the operation of quasi-autonomous driving vehicles (SAE level-3), but not for self-driving vehicles (SAE level-4). The legalization of fully autonomous vehicles is still being debated, mainly due to issues related to the determination of the rules for criminal and civil liability in the event of traffic accidents.[25]

The existing regulations of automated vehicles specify safety standards and mark the boundaries of the legalization of the levels of SAE automation, but leave questions open on how existing liability rules should be tailored to allocate accident losses. For example, the interaction of negligence torts and products liability is indeterminate when the driver of a vehicle is a robot. In an ordinary car accident, the human driver is liable under negligence torts if he/she failed to exercise due care, and the manufacturer is liable if the accident was caused by a manufacturing defect or design defect. If there is neither negligence nor a product defect, then the victim is left uncompensated for the accident loss.

On the one hand, it could be argued that robot torts fall within the domain of products liability because the self-driving software is simply part of the car. It is well established that automobile manufacturers have a duty to ensure that the design of an automobile mitigates danger in case of a collision (*Larsen v. General Motors Corp.* (391 F.2d 495 [1968])). This rule would naturally extend to self-driving cars, where manufacturers are afforded greater opportunity to avert or mitigate accidents, thereby expanding their duty of care. The standard for demonstrating a defect in a self-driving car can be inferred from existing case law. For example, in *In re Toyota Motor Corp. Unintended Acceleration Mktg., Sales Practices & Prods. Liab. Litig.* (978 F. Supp. 2d 1053 [2013]), vehicles produced by Toyota automatically accelerated without any driver action, and the plaintiffs were granted recovery.[26]

[24]See, e.g., "U.K. government announces Automated Lane Keeping System call for evidence." Published on August 18, 2020. Available at https://www.gov.uk/government/news/uk-government-announces-automated-lane-keeping-system-call-for-evidence. See also "Rules on safe use of automated vehicles on GB roads." Published on April 28, 2021. Available at https://www.gov.uk/government/consultations/safe-use-rules-for-automated-vehicles-av/rules-on-safe-use-of-automated-vehicles-on-gb-roads.

[25]See, e.g., "'Level 4' self-driving transit cars in Japan won't require licensed passengers: expert panel." Available at https://mainichi.jp/english/articles/20210402/p2a/00m/0na/025000c; and "Legalization of Self-Driving Vehicles in Japan: Progress Made, but Obstacles Remain." Available at: https://www.dlapiper.com/en/japan/insights/publications/2019/06/legalization-of-self-driving-vehicles-in-japan/. Last accessed: July 2021.

[26]See also *Cole v. Ford Motor, Co.* (900 P.2d 1059 [1995]) (holding the manufacturer liable when the cruise control

Similar reasoning could be transposed, *mutatis mutandis*, to self-driving vehicles.

On the other hand, it could be argued that robot torts fall within the domain of negligence torts, because autonomous driving is not qualitatively different from earlier innovations in automobile technology. Automation is not a discrete state, but rather a continuum. The electric starter, automatic transmission, power steering, cruise control, and anti-lock brakes have all increased the control gap between the operator and vehicle. Nonetheless, none of these technological innovations have excused the operator of tort liability. The move to autonomous driving will not be instantaneous, and it is unlikely to be total.[27] It is likely that for the foreseeable future operators will have the option to disengage autonomous operation. Indeed, it is plausible that there will be conditions where it would constitute negligence to engage autonomous operation.[28] As long as the operator is ultimately in control—even if that control only extends to whether autonomous operation is engaged or not—traditional tort doctrine identifies the *operator* rather than the manufacturer as the party that should be the primary bearer of liability.

Thus, reasonable arguments can be advanced for assigning liability to the manufacturer as well as the operator. However, claiming that robot torts should be adjudicated "in accordance with product liability law, common law, or other applicable federal or state law" merely begs the question.

Tort law is a blank slate with respect to self-driving cars. The *Federal Automated Vehicles Policy* merely suggests factors to consider when formulating a rule, whereas it does not recommend any particular liability rule. Indeed, the few states that have acknowledged the issue have merely booted the problem to be resolved by existing law, despite the existing law's indeterminacy on the novel question.

---

function caused the car to accelerate unexpectedly). See generally *Greenman v. Yuba Power Products* (59 Cal. 2d 57 [1963]); *Escola v. Coca-Cola Bottling Co.* (24 Cal. 2d 453 [1944]); *Ulmer v. Ford Motor Co.* (75 Wash. 2d 522 [1969]); *Restatement (Third) of Torts: Products Liability*.

[27] SAE Level 5 autonomous vehicles are defined as being able to drive under all conditions, but there may still be limits to certain components, e.g., heavy rain making it difficult to distinguish objects, under which an autonomous vehicle may not operate, but a human could.

[28] For instance, see Plumer, Brad (2016). "5 Big Challenges That Self-Driving Cars Still Have to Overcome." *Vox*. Available at https://www.vox.com/2016/4/21/11447838/self-driving-cars-challenges-obstacles (last updated April 21, 2016).

2.3.4 Medical Robots

Another recent and promising use of robot technology is in the field of medicine. Robots have been utilized in surgical operations since at least the 1980s, and their usage is now widespread (e.g., Lanfranco et al., 2004; Mingtsung and Wei, 2020).[29] Due to their better precision and smaller size, robots can reduce the invasiveness of surgery. Previously inoperable cases are now feasible, and recovery times have been shortened.

Some surgical robots require constant input from surgeons. For example, the da Vinci and Zues robotic surgical systems use robotic arms linked to a control system manipulated by the surgeon.[30] While the da Vinci and Zues systems still require input from a human operator, in other areas of medicine there is a general trend towards even greater robot autonomy. Many healthcare providers are beginning to use artificial intelligence to diagnose patients and propose treatment plans. These artificial intelligence systems analyze data, make decisions, and output results, although the results may be overridden by a human operator or supervisor (Kamensky, 2020). As the technology further develops, it is plausible that surgical robots will require even less input from operators.

The applicable tort regime for medical robots is still evolving. Allain (2012) provides an overview of the tort theories that victims have used in cases involving surgical robots, including medical malpractice, vicarious liability, products liability, and the learned intermediary doctrine. In instances where medical professionals actively control surgical robots, victims often assert medical malpractice claims that focus on the negligence of the medical professional, with reasonableness standards evolving over time based on advances in technology and knowledge. If the surgical robot or artificial intelligence is deemed a medical product—and therefore subject to Food and Drug Administration regulations—victims also often assert a products liability claim against manufacturers (Marchant and Tournas, 2019). However, this area of law remains relatively undefined, especially in cases involving software only.[31]

[29] It is also worth mentioning the spate of peer-reviewed scholarly journals dedicated to the topic that arose during this period. For example, the *International Journal of Medical Robotics and Computer Assisted Surgery*, established in 2004, the *Journal of Robotic Surgery*, established in 2007, the *American Journal of Robotic Surgery*, established in 2014, and the *Journal of Medical Robotics Research*, established in 2016.

[30] In a recent case involving the da Vinci robotic surgical system, a Florida man died as a result of a botched kidney surgery. The family claimed negligence based on the surgeon's lack of training and experience with the system, but the case was later settled out of court (Allain, 2012).

[31] In one recent decision involving a surgical robot, *Thomas v. Intuitive Surgical, Inc.* (389 P.3d 517 [2017]), the court

As with self-driving cars, victims currently have no clear liability regime under which to seek compensation from operators or manufacturers for autonomous medical robots. At present, fully autonomous medical robots are still relatively uncommon; however, machines are taking on an ever-increasing share of decision-making tasks (see Kassahun et al., 2016) The tort issues that have been litigated thus far have tended to revolve around operator error (see, for example, *Taylor v. Intuitive Surgical, Inc.* (389 P.3d 517 [2017])). Thus, for our purposes—much like self-driving car accidents—the law of medical robot torts is a *tabula rasa*.

2.3.5 Military Robots

Military drones and robotic weapons are another area where robot torts are implicated. These machines are already being used to identify and track military targets. Additionally, weaponized drones have been used extensively in lethal combat. The UN Security Council Report of March 8, 2021 (UN S/2021/229) regarding a Turkish military drone that autonomously hunted humans in Libya without any human input or supervision in March 2020 is just the first of possibly many instances of autonomous attacks by military robots. During recent years, media speculation about this topic has been rampant and the recent Libya incident has revived the debate.[32] It is easy to imagine other circumstances in the near future where constant communication with a human operator may not be possible and the identification and killing of an enemy target will be conducted autonomously. Should military technology continue to develop along this trajectory, it seems inevitable that other innocent targets will be attacked and eventually killed.

At present, no legal framework exists in the US to address a mistaken killing by a military robot. Regarding the civilian use of non-military drones, the Federal Aviation Administration has begun to

sought to decide whether the surgeon—i.e., the operator—or the manufacturer of the robotic surgical system would be liable and whether the manufacturer had a duty to warn. The court held that medical device manufacturers have a duty to warn hospitals and operators of the systems of the risks of the system.

[32]See, for example, Campaign to Stop Killer Robots, "Country Positions on Negotiating a Treaty to Ban and Restrict Killer Robots," (September 2020). Available at https://www.stopkillerrobots.org/wp-content/uploads/2020/05/KRC_CountryViews_25Sep2020.pdf; (last accessed August 13, 2021); Grothoff, Christian and J.M. Porup. 2016. The NSA's SKYNET Program May Be Killing Thousands of Innocent People. *Ars Technica*, February 16. Available at https://arstechnica.co.uk/security/2016/02/the-nsas-skynet-program-may-be-killing-thousands-of-innocent-people/; "A Military Drone With A Mind Of Its Own Was Used In Combat, U.N. Says" (June 1, 2021). Available at https://www.npr.org/2021/06/01/1002196245/a-u-n-report-suggests-libya-saw-the-first-battlefield-killing-by-an-autonomous-d (last accessed August 13, 2021).

address ways to regulate drone usage within the US in recent years, although it has not yet systematically addressed liability for physical harm (Hubbard, 2014).[33] In an August 2021 Report released by the Human Rights Watch and the Harvard Law School International Human Rights Clinic, a proposal has been presented for a normative and operational framework on robotic weapons. States that favored an international treaty regulation of autonomous weapon systems agreed that humans must be required to play a role in the use of force, with a prohibition of robotic weapons that make life-and-death decisions without meaningful human control..[34]

2.3.6 Other Uses

Robots are also used in factories and other industrial settings due to their ability to quickly and efficiently execute repetitive tasks. When an industrial robot injures a victim, it often occurs in the context of employment. In such instances, workers are typically limited to claiming workers' compensation and barred from asserting tort claims against their employer. Many states include exceptions to this rule for situations where the employer acted with an intent to injure or with a "deliberate intention" of exposing the worker to risk. However, thus far most of the cases brought by victims have proven unsuccessful (Hubbard, 2014). Due to the relatively controlled environment of factories and other industrial settings, operators can typically ensure a relatively high probability of safe operation and prevent injuries to potential victims.

3 A Theory of Manufacturer Residual Liability

3.1 Terminology and Scope

The scenarios that we seek to analyze to delimit the scope of our inquiry arise from the interaction of three parties: an operator, a victim, and a manufacturer. The manufacturer develops and/or produces

[33] For a review of unmanned aerial vehicles regulations on the global scale, see Stöcker et al., 2017 and Jones, 2017; for European cases, see Esteve and Domènech, 2017; for military drone laws and policies in Japan, see Sheets, 2018.

[34] See the Report "Areas of Alignment: Common Visions for a Killer Robots Treaty" which presents the objections expressed by governments at the official Convention on Conventional Weapons (held in September 2020) to delegating life-and-death decisions to robots. Available at https://www.hrw.org/sites/default/files/media_2021/07/07.2021%20Areas%20of%20Alignment.pdf (last accessed August 12, 2021).

the robot. The operator utilizes the robot to carry out an activity, the actions of which can cause harm to a victim. Ultimately, our goal is to determine what liability rule efficiently allocates the accident loss in these circumstances.

We use the term "operator" to refer to the human who gives a robot its objectives. In order to study the effect of alternative liability rules on these parties' incentives, in our analysis we always assume the existence of a human operator who is capable of directing the activity of the robot and overriding its decisions. We thus exclude from our analysis "humans-out-of-the-loop" robots that are completely capable of self-determination.[35]

Although a robot is a decision-making machine by definition, it cannot decide its own objectives; rather, it must receive its objectives from a human. For example, a self-driving car is such a decision-making machine. It must make thousands of choices to navigate to a destination, locate a parking spot, or park. However, the decision *to go to the restaurant*, *circle the block*, or *park* must ultimately come from a human. Given an objective, the robot decides the *means* of accomplishing those ends, but a human still decides what the *ends* are. We use the term "victim" to refer to an individual who has suffered harm as a result of a robot's choices. We are principally interested in circumstances where the operator and victim are distinct individuals. However, we show that the special case where the operator is also the victim reduces to an ordinary case of products liability.

We use the term "manufacturer" to refer to an entity responsible for the development, production, and sale of a robot. There may be several entities that satisfy the definition of "manufacturer" in a given case. Our analysis seeks to answer only whether (and how much) liability should be assigned to a monolithic "manufacturer." The allocation of liability between multiple manufacturers is a further question beyond the scope of our inquiry.

Next, we distinguish between torts *caused by a robot's decisions* versus torts that merely *happen to involve a robot*. For example, if a robot surgeon excises healthy tissue, mistaking it for a tumor, it is the robot's *decision* that causes the harm. On the other hand, if an automobile crashes into a pedestrian

[35]As pointed out by Casey (2019), Bryan Casey, Robot Ipsa Loquitur, 108 Geo. L.J. 225 (2019), although the level of autonomy of a robot exists on a spectrum, its immediate legal consequences are binary: the actions that led to the accidents either involved tasks where robots operated with human intervention—referred to as "humans-in-the-loop" activities—or tasks that neither involved nor needed human intervention, referred to as "humans-out-of-the-loop" activities. With fully autonomous robots, operators would have no control over the robot and their incentives would no longer be a relevant focus in the analysis

because the brakes malfunctioned*, it is immaterial whether the car was being driven by a robot or a human, because the accident was not the consequence of the robot's decision. In cases where defects in other (non-decision-making) components of the machine cause harm, the problem falls within the standard realm of products liability law. The theoretically interesting cases that are the focus of our paper are those where the robot's *choices* cause harm.

We also need to distinguish what it means for a robot to cause an "accident." While by definition an accident is an *unintended harm*, robots do not have "intentions" in the ordinary sense. It could therefore be argued that *all* of a robot's choices are accidents, or that *none* of a robot's choices are accidents. A robot merely executes an algorithm, a fixed set of instructions. Insofar as the robot executes the algorithm, it has "succeeded" in what it was meant to do. The sense in which it has caused an "accident" is that the result was not intended or caused by the negligence of the robot's operator. This is what will mark the boundary between an operator's fault-based liability and a manufacturer's strict residual liability in our proposed rule.

To determine the optimal tort liability regime for robots torts, we first begin by noting several features of robot torts that are distinguishable from human torts. First, robot decision-making is most often a *substitute* for human decision-making. Unlike conventional machines, robots do not merely enhance the efficacy of human actors, but rather they replace human actors. The substitution is not necessarily total, especially in the initial stages. A robot may be partially autonomous, sharing decision-making duties with a human operator, or a robot may be fully autonomous but allows the operator to veto its decisions or override its operation altogether.

Second, the concept of negligence cannot be meaningfully applied to robots. Robots do not exercise precautionary care, but rather they mechanically execute an algorithm. The "effort" required for a robot to execute a "reasonable algorithm" is no greater than the "effort" required to execute an "unreasonable algorithm." Thus, the probability of an accident arising due to machine error is not a function of the robot's "care level." Rather, it is inherent in the quality of the robot's algorithm, which is a function of the manufacturer's research investment, the research invested in developing safer robots functions as a *durable precaution*. Once a safer technology is developed, there is no meaningful risk that it will subsequently be "forgotten."

As discussed above, in robot torts, robots carry out activities that are partially controlled by a human operator. In turn, these activities can cause harm to third-party victims. The care of the operator as well as the prospective victim can often reduce the likelihood of an accident. The problem for a policymaker is to design liability regimes that incentivize both the operator's and victim's optimal care, while at the same time inducing manufacturers to internalize the expected cost of non-negligent accidents and improve the safety of the robots.

Past literature has principally focused on product malfunctions, where the victim is the product user or a third party. The general approach used in this paper of analyzing robot-generated accidents enables considering cases not necessarily characterized by a design flaw or product defect. The analysis starts from the premise that an accident arising from the actions of a robot might not be, *ipso facto*, the result of a design flaw or manufacturing defect. A perfect and risk-free functioning of a robot is an unreachable ideal, and accidents caused by a robot cannot *per se* be construed as a product defect. To better see this point, consider the case of self-driving automobiles. Very specialized circumstances may arise when a human driver would avoid an accident better than a robot. Because an automated technology operates by applying general rules, there may always be special cases that fall outside the scope of the algorithm where a human decision-maker might perform better. This is what programmers call "corner cases," namely situations not defined by any well-understood rules. Nonetheless, failing to account for every special circumstance—which are innumerable in number—should not be regarded as a design flaw or error, particularly if robots outperform humans in the vast majority of "normal" situations.

In this paper, we show that a rule of manufacturer residual liability outperforms rules of liability for product malfunction. We refer to this class of liability rules generally as "manufacturer residual liability." These regimes can accomplish multiple objectives. This class of regimes assigns primary liability on the operator and/or victim, contingent upon their negligence.

In Figure 1, we illustrate the allocation of liability under three negligence rules that can be coupled with the manufacturer's residual liability: (a) simple negligence; (b) contributory negligence; and (c) comparative negligence. As shown in Figure 1, under all three rules, if the operator is the only negligent party (bottom-left quadrant in each matrix), the operator bears the accident loss, and if the

		Operator	
		Negligent	**Diligent**
Victim	**Negligent**	Operator	Victim
	Diligent	Operator	Manufacturer

(a) Simple Negligence with Manufacturer Residual Liability

		Operator	
		Negligent	**Diligent**
Victim	**Negligent**	Victim	Victim
	Diligent	Operator	Manufacturer

(b) Contributory Negligence with Manufacturer Residual Liability

		Operator	
		Negligent	**Diligent**
Victim	**Negligent**	Share	Victim
	Diligent	Operator	Manufacturer

(c) Comparative Negligence with Manufacturer Residual Liability

Figure 1: Manufacturer Residual Liability Rules

victim is the only negligent party (top-right quadrant in each matrix), the victim bears the accident loss. Regardless of the primary liability regime where both the operator and victim are diligent (bottom-right quadrant in each matrix), our rules of manufacturer residual liability shift the accident loss onto manufacturers.

The only difference between the three rules is the way in which they allocate the accident loss when the operator and victim are both negligent (top-left quadrant in each matrix). Under a simple negligence rule, the focus is exclusively on the behavior of the operator: a negligent operator is liable to compensate the victim, regardless of the victim's behavior.[36] Under a contributory negligence rule, a negligent victim would be completely barred from obtaining compensation in a negligence case.[37] Under a comparative negligence rule, when both operators and victims are at least somewhat

[36]Simple negligence rules are easier to administer, because courts only need to evaluate the behavior of tortfeasors. They are adopted in cases where victims have little control over the risk of accidents (i.e., unilateral care cases).

[37]Contributory negligence rules are generally used in products liability cases. Misuse of a product or a plaintiff's failure to follow clear instructions and/or warnings provided by the manufacturer are often construed as contributory negligence. Paraphrasing Weston (1963), under a contributory negligence rule, victims of a robot would not be permitted to take refuge from their own stupidity due to a breach of a duty of care by operators.

negligent, the accident loss would be divided between them as a percentage based on their relative degree of negligence.[38]

As will be shown in Section 3.4, the different allocations of liability in cases of bilateral negligence do not affect the parties' care and activity-level incentives. Therefore, in our discussion we shall jointly discuss the incentive effects of all three manufacturer residual liability rules presented above.

3.2 A Basic Model

Let us begin with definitions. We denote the human operator by O, the victim by V, and the robot manufacturer by M. We represent the operator's precautionary care level by x and the operator's activity level by w. The value of the operator's activity is described by $V_O(w)$. We assume diminishing marginal benefits $V_O' > 0, V_O'' \leq 0$. Next, we represent the victim's investment in precautionary care level by y and the victim's activity level by z. The value of the victim's activity is described by $V_V(z)$. Again, we assume diminishing marginal benefits $V_V' > 0, V_V'' \leq 0$. Without loss of generality, we will not explicitly consider the availability of robot technology for the victim.[39]

Finally, we represent the manufacturer's investments to produce the robot by r. Production costs include R&D investments for the development and improvements in the quality and safety of the robots. The accident loss is represented by $L > 0$, and the probability of an accident occurring is represented by $p(x,y,r)$. We assume that precautionary care and investments in researching robot technology reduce the probability of accidents, i.e., that $\frac{\partial p}{\partial x} < 0$, $\frac{\partial p}{\partial y} < 0$, and $\frac{\partial p}{\partial r} < 0$. We also assume the diminishing marginal effectiveness of precautions and research investments, i.e., that $\frac{\partial^2 p}{\partial x^2} \geq 0$, $\frac{\partial^2 p}{\partial y^2} \geq 0$, and $\frac{\partial^2 p}{\partial r^2} \geq 0$. Finally, we assume that investments in r reduce the effectiveness of human precautionary care, i.e., that $\frac{\partial^2 p}{\partial x \partial r} \geq 0$, $\frac{\partial^2 p}{\partial y \partial r} \geq 0$.

At this point, there are several ways to model robot torts. The most literal approach would be to define the social cost function by $wzp(x,y,r)L + wx + zy + r$. However, in the interest of simplicity, we

[38] Several jurisdictions in the US have abandoned the contributory negligence rule in favor of comparative negligence rules in ordinary negligence cases (non-products liability), when bilateral negligence is established, since the former rule would bar victims from obtaining compensation, even when the negligence of the tortfeasor is much more serious. Most civil law systems also utilize comparative negligence rules in apportioning damages in bilateral negligence cases.

[39] As the probability of symmetrical use arises (where both tortfeasors and victims use robots), the efficient standard of care will tend to be greater than in the unilateral use case (where only the injurer is using a robot). This follows directly from Luppi et al. (2016).

instead define the expected accident loss by $\frac{wz}{r}p(x,y)L$ without loss of generality. This representation is more elegant and requires fewer assumptions. We need to only assume of the probability function that $\frac{\partial p}{\partial x} < 0$, $\frac{\partial p}{\partial y} < 0$, $\frac{\partial^2 p}{\partial x^2} \geq 0$, and $\frac{\partial^2 p}{\partial y^2} \geq 0$. This faithfully represents the general characteristics of the robot tort problem: (1) the term r (without an activity-level multiplier) captures the fact that research investments are "durable precautions," (2) the assumptions $\frac{\partial^2 p}{\partial x \partial r} \geq 0$ and $\frac{\partial^2 p}{\partial y \partial r} \geq 0$ capture the fact that automation does not supplement but rather *replaces* human precautionary care, and (3) the assumption that $\frac{\partial p}{\partial r} < 0$ captures the fact that increasing investments in researching robot technology tends to reduce the expected accident loss.

3.3 Social Optimization Problem

To determine how the various tort liability regimes alter the care and activity levels of the operator, victim, and manufacturer, let us define the social welfare function. In our three-party setup, the social welfare function takes the following form:

$$\max_{x,y,w,z,r} S = V_O(w) + V_V(z) - \frac{wz}{r}p(x,y)L - wx - zy - r. \tag{3.1}$$

The efficient levels of the operator's care, x, and victim's care, y, are described by the following first-order conditions:

$$x^{**} : -\frac{wz}{r}\frac{\partial p}{\partial x}L = w \tag{3.2}$$

$$y^{**} : -\frac{wz}{r}\frac{\partial p}{\partial y}L = z. \tag{3.3}$$

Care investments are efficient when the marginal reduction in the expected accident loss equals the marginal cost of care.

The efficient levels of the parties' activity levels, w and z respectively, are described by the following first-order conditions:

$$w^{**} : V_O' = \frac{z}{r} pL + x \tag{3.4}$$

$$z^{**} : V_V' = \frac{w}{r} pL + y. \tag{3.5}$$

Activity levels are efficient when the marginal benefit from an increase in activity level equals the marginal cost of the activity.

Finally, the efficient research investment, r, is described by the following first-order condition:

$$r^{**} : \frac{wz}{r^2} pL = 1. \tag{3.6}$$

The manufacturer's research investment is efficient when the marginal benefit is equal to the marginal cost.

The social welfare functions determine the efficient behavior of the parties. To determine the actual incentives of the parties, we must next analyze the private welfare functions.

3.4 Private Optimization Problem

We earlier identified three species of manufacturer residual liability rules, including simple negligence, contributory negligence, and comparative negligence (Figure 1). For ease of exposition, we introduce the variable θ to represent the share of the accident loss that the injurer must bear. Under simple negligence with manufacturer residual liability, $\theta = 1$; under contributory negligence with manufacturer residual liability, $\theta = 0$; and under comparative negligence with manufacturer residual liability, $0 < \theta < 1$. We can now generalize across the species of manufacturer strict liability. The private welfare functions of the operator, victim, and manufacturer, respectively, are:

$$\max_{x,w} U_O = \begin{cases} V_O(w) - wx & \text{if } \{x \geq x^{**} \wedge y < y^{**}\} \vee \{x \geq x^{**} \wedge y \geq y^{**}\} \\ V_O(w) - \theta \frac{wz}{r} pL - wx & \text{if } x < x^{**} \wedge y < y^{**} \\ V_O(w) - \frac{wz}{r} pL - wx & \text{if } x < x^{**} \wedge y \geq y^{**} \end{cases} \tag{3.7}$$

$$\max_{y,z} U_V = \begin{cases} V_V(z) - zy & \text{if } \{x < x^{**} \wedge\ y \geq y^{**}\} \vee \{x \geq x^{**} \wedge\ y \geq y^{**}\} \\ V_V(z) - (1-\theta)\frac{wz}{r} pL - zy & \text{if } x < x^{**} \wedge\ y < y^{**} \\ V_V(z) - \frac{wz}{r} pL - zy & \text{if } x \geq x^{**} \wedge\ y < y^{**} \end{cases} \tag{3.8}$$

$$\min_{r} U_M = \begin{cases} r & \text{if } x < x^{**} \vee\ y < y^{**} \\ r + \frac{wz}{r} pL & \text{if } x \geq x^{**} \wedge\ y \geq y^{**}. \end{cases} \tag{3.9}$$

Solving the private optimization problem, under a rule of manufacturer residual liability, combined with any of the three negligence-based regimes under consideration (i.e., simple negligence, contributory negligence and comparative negligence, hereafter jointly referred to as "rules of manufacturer residual liability"), we obtain the following results:[40]

Proposition 3.1. *Rules of manufacturer residual liability create optimal care incentives for robot operators and their prospective victims, and create optimal incentives for the manufacturer's R&D investments.*

Proof. See the Appendix. □

From Equations (3.7), (3.8), and (3.9) above, we see that all three rules of manufacturer residual liability create optimal care incentives on operators and victims and optimal R&D investments for manufacturers. These incentives will lead to bilaterally careful behavior and research in the development of safer robots.

Corollary 3.2. *Rules of manufacturer residual liability will lead to the production and maintenance of safer robots. In a competitive market, the cheapest robots will be optimally safe.*

As will be more extensively discussed in Section 4.1.1, under manufacturer residual liability, robots will be produced and maintained at an optimally safe level. This result does not hinge on the observability of the robots' quality by consumers at the time of the sale, nor the consumers' willingness to pay a premium for safer robots, as required in Hay and Spier (2005, p.1701). Instead,

[40] Consistent with convention, these optimality results are in fact second-best results, inasmuch there are second-order effects of excessive activity levels on care expenditures. See Dari-Mattiacci et al. (2014) and Carbonara et al. (2016) for a further discussion of these effects in negligence regimes.

this simply follows from considering that the manufacturer's R&D investments are durable, i.e., once the safer algorithm has been developed, no additional investment is required to maintain that level of safety. More specifically, under rules of manufacturer residual liability, manufacturers will have incentives to minimize their expected cost K also by including in r the cost of updates and maintenance of the automated technology. Their total expected investment in r—which includes their expected investment in post-sale safety research—will be up to the point where $\frac{wz}{r^2}pL = 1$, which is equal to the socially optimal level as defined in Equation (3.6). The manufacturer invests in safety research until the last dollar spent reduces expected injury costs by one dollar, which is characterized by the socially optimal level of r. By adopting these socially optimal levels of investments in R&D, the manufacturer will develop and maintain robots at an optimally safe level, and the safer robots will prove to be the more affordable ones.

In turn, the manufacturer's higher R&D investment helps in minimizing the robot's price. More formally, the manufacturer faces a total production cost, K, which includes *both* the fixed R&D investments for safety and quality, r, *and* the expected residual liability costs, $\frac{wz}{r}pL$. In a perfectly competitive market, the manufacturer maximizes profits by setting the price equal to the marginal cost, i.e., $\frac{wzpL}{r^2}$. This implies that safer robots will be less expensive, while more dangerous robots will be more expensive (as r increases, the marginal cost—hence, the price—decreases), further generating optimal incentives for the production of safer robots.

Corollary 3.3. *Rules of manufacturer residual liability create optimal incentives for the gradual adoption of safer robots.*

As will be more extensively discussed in Section 4.1.2, all rules of manufacturer residual liability create optimal incentives for operators that already possess robots to gradually upgrade technology and adopt safer robots. The investment in new robots will be initially undertaken by high-activity-level operators, and later by others. This gradual adoption mechanism will have interesting allocative efficiency properties, matching newer and safer robots with operators that plan to make greater use of the robot. This simply follows from the fact that investments in r reduce the effectiveness of human precautionary care, i.e., that $\frac{\partial^2 p}{\partial x \partial r} \geq 0$, $\frac{\partial^2 p}{\partial y \partial r} \geq 0$. When using a very safe robot, the socially optimal levels of human care as defined in Equations (3.2) and (3.3) decrease. This means that adopting safer technology reduces individuals' marginal cost of an additional unit of activity. Hence, by adopting safer robots, operators with high activity levels have greater savings in care costs than operators with low activity levels. Allocative efficiency will result, whereby older robots will initially be replaced with newer, safer robots by those who use them the most.

Proposition 3.4. *In the absence of a price relationship between the parties, rules of manufacturer residual liability cause excessive activity levels for both operators and victims.*

Proof. See the Appendix. □

From Equations (3.7) and (3.8), we can see that the allocation of residual liability on manufacturers leads to excessive activity levels for both operators and victims, since neither of these parties expects to face any liability in equilibrium. The misalignment of incentives occurs because operators (and victims) derive benefits from the use of robots (and activities that expose them to robots), but an increase in their activity level increases the probability of accidents, with a resulting externality on the manufacturers' residual liability.

Corollary 3.5. *When manufacturers can measure the robot's usage and charge the cost of residual liability to operators, optimal activity levels will be undertaken.*

As will be more extensively discussed in Section 4.2, with current technology, robots can keep track of the operator's activity level (e.g., keeping track of the mileage of self-driving cars or the number of surgeries performed by robots). In a competitive market, manufacturers would have strong incentives to develop a price mechanism to transfer the marginal cost of the risk created by these activities back to operators. This pricing mechanism will induce operators to internalize the risk that they create on manufacturers, and in turn their activity levels will converge to socially optimal levels.

For a simple, formal description of the mechanism stated in Corollary 3.5, we can refine the operator's private optimization problem as in the following:

$$max_w U_O = V_O(w) - w(\tau + x^*) \tag{3.10}$$

where τ is the fee charged by the manufacturer to the operator per activity level (e.g., fee per mileage), and $x^* = x^{**}$ as defined in (3.2). The manufacturer will set τ equal to the cost of residual liability per operator's activity level, i.e., $\tau = \frac{z}{r}pL$. The operator's privately optimal activity level, $\bar{w}^*$, becomes:

$$\bar{w}^* : \; V_O' = \frac{z}{r}pL + x^* \tag{3.11}$$

It follows that $\bar{w}^*$ converges to the socially optimal level w^{**} as defined in Equation (3.4). Additionally, if operators and victims are contractually related (e.g., the operator is using the robot to offer a service to the victim), the cost of the service would increase to reflect the extra fees charged for the use of the robot by the manufacturer to the operator. As a result of this price increase, incentives would percolate from manufacturers to victims, and the activity level of the victim would also be mitigated.

In Section 4, we provide a more extensive discussion of these Propositions and Corollaries and other interrelated effects of our manufacturer residual liability rules.

4 Effects of Manufacturers' Residual Liability

4.1 The Market for Robots

In the foregoing discussion, we suggested that in most situations the optimal liability regime for robot torts is one where the manufacturer is the sole residual bearer of the accident loss. The logic behind this idea is two-fold. First, when manufacturers face residual liability, they will have optimal incentives to innovate and improve their products, especially considering that the most technologically complex parts of robots are more prone to undetectable failures. When the costs of developing safer technologies are not verifiable in court, incentivizing the creation of safer robots through negligence rules is unlikely to be a successful policy lever. When the costs and availability of safer technology are not verifiable in court, the production of safer robots cannot be easily incentivized by the threat of negligence liability. Notwithstanding any attempt to modify negligence standards, it would be difficult to hold manufacturers negligent and impose liability on them for not having developed safer robots.[41] Courts have no direct information to establish what would be the socially optimal advances in technology, and those decisions are best delegated to manufacturers, who have direct information

[41] However, see Dari-Mattiacci and Franzoni (2014), suggesting the possibility that negligence standards could be adjusted upwards or downwards when adoption costs are not verifiable in court, depending on whether the adopted technology reduces or increases expected harm.

about the costs and benefits of technological safety. Second, we should incentivize operators to adopt the safer robot technologies that manufacturers develop. A desirable liability rule therefore needs to incentivize both manufacturers and operators to make investments in safer robots. Incentives to invest in more advanced and safer robot technologies generally fall on the party that bears residual liability.

As shown in Proposition 3.4, under our rule, manufacturers face the threat of residual liability and thus have incentives to invest in R&D to produce safer robots. Because residual liability can only be placed on one party, it may seem that we cannot simultaneously incentivize both manufacturers to produce safer robots and operators to adopt them.[42] However, as discussed in Sections 4.1.1 and 4.1.2, our proposed manufacturer liability rule can overcome this difficulty, creating incentives for manufacturers to produce safer and cheaper robots, and creating incentives for operators to purchase them.

4.1.1 Manufacturers' Research Incentives and the Pricing of Safer Robots

In many products liability models, the belief that safer products will develop in the market rests on two fundamental assumptions: (i) that consumers are willing to pay a premium for safer products, and (ii) that product safety is perfectly observable to consumers at the time when they make their purchasing decisions (Hay and Spier, 2005). However, under our manufacturer residual liability rule, neither of these assumptions is likely to hold.

First, in the three-party scenario under consideration, robot operators are only interested in avoiding liability, which they can do by adopting due care in the handling of the robot. They would not be willing to pay a premium to acquire a safer robot, because investments in safety would reduce the risk of accidents, not their expected liability. Second, robot technology is relatively complex. Problems and shortfalls generally materialize in the course of the robot's operation: in other words, a robot's safety is not observable by operators at the time of their purchasing decision. The specific design limitations manifest themselves over time through the use of the robot, and they are unknown to the operator, just as they are often unknown to the manufacturer at the time of production.

[42]Most readers will recognize this problem as a three-party incarnation of Shavell's (1980) activity-level theorem. Shavell's theorem holds that only the bearer of residual liability is incentivized to undertake precautions that are not incentivized by the negligence standard. This is because the party who does not bear residual liability only wants to avoid liability by showing that he adopted due care, whereas the bearer of residual liability wants to avoid causing harm *tout court*.

However, as shown in Corollary 3.2, when manufacturers face residual liability, they fully internalize the benefits of safety of their products. Once the robot is in the hands of the operator, the manufacturer is unable to influence the risk of injury, and all non-negligent harm caused by the robot imposes an expected cost of liability on the manufacturer equal to $\frac{wz}{r}pL$. As discussed under Corollary 3.2, the expected cost of future non-negligent accidents becomes part of the cost of the product. When determining their optimal total investment in quality and technology—which includes the ex-ante investments in R&D for developing the robot and post-sale safety updates to improve the algorithm of existing robots—they will balance safety investments and expected liability costs. These optimal total costs will determine the price of their product in a competitive market.

By investing in development and post-sale R&D, the manufacturer affects the level of risk that the robot causes with its operation by its operators and its potential future liability. Corollary 3.2 shows that in a legal regime where operators are only liable for negligent accidents and search for robots at the lowest price, manufacturers will have incentives to minimize the cost K and the sales price of the automated technology, and therefore invest in the safety and quality r up to the point where $\frac{wz}{r^2}pL = 1$, which is equal to the socially optimal level as defined in Equation (3.6). The manufacturer invests in safety research until the last dollar spent reduces expected injury costs by one dollar, which is characterized by the socially optimal level of r. The resulting automated technology under a manufacturer residual liability rule is therefore optimally safe.

By adopting these socially optimal levels of investments in research, the manufacturer will develop and maintain robots at an optimally safe level, and the safer robots will prove to be the more affordable ones. Making the manufacturer internalize the full cost of the harm caused by the robot results in the robot's price reflecting the manufacturer's liability. Consequently, manufacturers will make their production decisions based on the total cost that they face $K = r + \frac{wz}{r}pL$, rather than looking at the bare development cost, r, that they would face in the absence of residual liability.

Under all three rules of manufacturer residual liability, the price of the robot would reflect its dangerousness, whereby more dangerous robots would be more expensive, and safer robots would be less expensive. Manufacturers would compete on price to sell their robots, and by doing so they would compete on safety, producing and maintaining robots with the socially optimal amount of safety, minimizing price. Even if operators are not held residually liable for the harm caused by the robot, competitive market forces would lead to the development and adoption of cheaper and safer robots, regardless of whether robot operators are informed about safety when making their purchasing decisions.

4.1.2 Operators' Adoption of Newer, Safer Robots

As shown by Corollary 3.3, rules of manufacturer residual liability create incentives for high-activity-level operators that already possess robots to upgrade technology and adopt safer robots. The gradual spread of new robots in the market has interesting allocative efficiency properties, since it allocates newer and safer robots in the hands of operators that plan to make greater use of them. The explanation for this gradual adoption of newer technology is given by the fact that an increase in the safety of a robot decreases the need for—and effectiveness of—human precautionary care. Newer, safer robots will therefore hold greater value to those who plan to use them more.

However, will low-activity-level operators be incentivized to upgrade obsolete, and unsafe, robots? Notwithstanding the lower pricing of newer and safer robots introduced in the market through the mechanisms described in Section 4.1.1, some existing owners may continue to use older robots. Some of these robots may become relatively more dangerous as safety standards improve. In these situations, there exist several indirect mechanisms to induce operators to adopt safer upgrades.

Let us first consider possible market solutions. Under a regime of manufacturer residual liability, manufacturers of older robots face higher levels of expected liability. Manufacturers will therefore be strongly incentivized to find ways to upgrade or replace their obsolete robots. There are several mechanisms by which this may be accomplished. Manufacturers could offer maintenance plans or they could provide free firmware or hardware upgrades to improve the safety of legacy robots.

Alternatively, anticipating the higher risk created by aging robots, manufacturers could set expiration dates for the usability of their robots, or they could adopt a leasing rather than sales model. A leasing model would provide manufacturers the option to replace older robots with upgraded models upon the renewal of each term of the lease. The point is that the threat of residual liability creates a strong incentive for manufacturers to devise ways to remove and replace obsolete robots from use and encourage operators to adopt and use the safest robot technologies.

Nonetheless, even if manufacturers are incentivized to aggressively control which of their products remain in use, there could still be situations where obsolete and unsafe robots remain in operation. For example, a manufacturer that has gone out of business will not be responsive to threats of tort liability. In these cases, other legal instruments may need to be utilized. The first and most obvious solution would simply be to construe the operator's use of an obsolete robot as per se negligence (see Dari-Mattiacci and Franzoni, 2014). By designating the use of an obsolete robot as presumptively negligent conduct, the primary care incentives of the operator would be activated.

However, this may still leave some cases uncovered. For example, when the operator is unaware of the age of the robot, primary care incentives to upgrade may be rendered ineffective. In this case, obsolete robots may be replaced or removed by direct regulation. For example, robots may be required to undergo periodic safety inspections. In cases where a dangerous obsolete robot is observed, the inspection authority can simply decertify its use.

However, overall our proposed manufacturer residual liability rule achieves the objectives of a desirable liability regime by incentivizing the care level of both operators and victims, as well as incentivizing manufacturer investments in safer robot technologies. Even though the rule cannot directly incentivize operators to adopt these safer robot technologies, manufacturers will be incentivized to better control their legacy technologies and encourage the quicker adoption of newer and safer ones. Manufacturers are in a better position than operators to control the safety of robots, and therefore they should be assigned sole residual liability.

4.2 Percolation Effects on Activity Levels

Regardless of which species of manufacturer residual liability regime is chosen, negligence rules cannot incentivize non-verifiable precautions because non-verifiable precautions are—by definition—undetectable by courts in determining negligence.[43] In traditional accident cases involving an injurer

[43] The allocation of residual liability is a policy lever that can be used to generate different subsets of "non-verifiable precaution" incentives on the various parties. Shavell (1980) called the measure of non-verifiable precautions the "activity level" of the parties. The rate at which a party undertakes an activity is often only one among a variety of possible non-verifiable precautions. Non-verifiable precautions include not only reductions in activity levels, or looking in the rear-view mirror while driving a car, but also investments in research for making activities safer. From an economic point of view, in the standard injurer/victim problem, residual liability should be imposed on the party whose non-verifiable precautions most effectively reduce the cost of accidents (Carbonara et al., 2016).

and a victim, the creation of incentives for "non-verifiable precautions" is accomplished through the allocation of residual liability. In our three-party scenario with operators, victims, and manufacturers, different parties have control over different aspects of non-verifiable precautions. Operators control the activity of the robots, manufacturers control research investments in robot safety, and victims control their own activity levels and non-observable precautionary efforts.

By allocating the residual liability to manufacturers, our rule only creates direct incentives for optimal "non-verifiable" R&D investments for manufacturers, leading to safer robot technologies.[44] This result was reflected in Proposition 3.4, where we observed that allocating residual liability on manufacturers may lead to excessive activity levels for both operators and victims, since these parties do not expect to face any liability in equilibrium. From a policy point of view, in the absence of a price mechanism, percolation effects would not emerge, and a trade-off would arise: incentivizing optimal activity levels by allocating residual liability on operators, or incentivizing optimal R&D by allocating residual liability on manufacturers.[45]

Corollary 3.5 shows that in the presence of price mechanisms (where manufacturers could charge operators a fee equal to the risk created by their activity level), the allocation of residual liability on manufacturers will likely percolate, leading operators and possibly victims to mitigate their activity levels. When liability rules are altered, markets will react. Under the proposed manufacturer residual liability rule, faced with the residual liability for robot accidents, manufacturers are likely to implement technical and contractual solutions to transmit at least some of the cost of their residual liability downstream to robot users (both operators and potential victims). This will then incentivize both operators and potential victims towards optimal activity levels.

Several mechanisms could be implemented to transmit incentives from manufacturers to operators through the price system. For example, robots could be designed to keep a record of usage rates, and manufacturers could make this information retrievable to monitor the operators' level of activity.[46] In

[44] Here, again we can see a three-party occurrence of Shavell's activity-level theorem (Shavell, 1980).

[45] As pointed out by one of our anonymous reviewers, we could think of an alternative rule of "operator's strict liability with manufacturer's contributory negligence." However, under this rule it would be difficult to fully incentivize manufacturers to improve the design of their robots with a threat of contributory negligence liability. Manufacturers would not fully internalize all the prospective benefit of their potential R&D, and it would be difficult to think of a price mechanism that could create an upstream percolation effect (from operators back to manufacturers) that could incentivize manufacturers to optimally improve the design of their robots.

[46] Unlike ordinary human activities, when robots are utilized, human activities are directly implemented and carried out

a competitive market, we should expect pricing mechanisms to shift the expected cost of non-negligent accidents associated with higher activity levels to operators. The expected cost of liability that robot activity levels create on manufacturers would therefore be passed on to the operators through the price system. In this way, the residual liability incentives faced by the manufacturers would percolate downstream to the operators, incentivizing them to engage in lower, socially optimal activity levels. Percolating some of the costs downstream will help to mitigate any excessive activity levels of the operators and potential victims and incentivize them to engage in socially optimal levels. If operators and victims are contractually related (e.g., the operator is using a self-driven car as a taxi, or the robot to offer a service to the victim), the cost of the liability would be included in the cost of the service, and the activity level of the victim would also be mitigated.[47] As discussed in Corollary 3.3, this will in turn incentivize high-usage operators to purchase safer technology that would impose a lower activity-level operation cost.

The proposed manufacturer residual liability rule is preferable to alternative allocations of residual liability. If operators were assigned residual liability, the inverse percolation of incentives would not be possible. While contractual and market mechanisms can easily be imagined to transmit residual liability incentives from robot manufacturers to robot operators under the manufacturer residual liability rule, no inverse mechanisms can be constructed as easily to transmit incentives upstream from robot operators to manufacturers.[48] Thus, to the extent that we want our assignment of residual liability to affect the incentives of as many parties as possible, assigning residual liability to the manufacturer is

by the robot. Activity levels can effectively be measured and recorded by the machine.

[47] In the absence of a price relationship between operators and victims, things may play out somewhat differently for the incentives of potential victims. If the manufacturer bears the entire residual liability, the potential victim may undertake excessive activity levels. For example, consider an accident between a self-driving car (the injurer) and a non-self-driving car (the victim): if the manufacturer of the self-driving car is held residually strictly liable in case of an accident, the driver of the non-self-driving car will engage in moral hazard and have reduced incentives to mitigate activity levels. The activity level of the victim cannot be as easily monitored by the robot and/or corrected through market or contractual mechanisms. However, faced with residual liability, manufacturers may have incentives to adopt technology capable of observing the behavior of victims that could render verifiable some of their—otherwise unobservable—precautions. Nevertheless, when the victims' activity levels are an overriding factor affecting the risk of non-negligent accidents, a different allocation of residual liability could be considered.

[48] Market mechanisms could hypothetically develop to transmit residual liability incentives upstream to manufacturers. For example, manufacturers could sell firmware or hardware upgrades that increase the safety of the robot, or operators could require manufacturers to provide an insurance coverage for liability arising from the sub-optimal quality of the robot. However, the greater degree of opacity of quality information compared to the easier measurability of the robot's activity level would probably render the upstream transmission of incentives more difficult in implementation compared to the downstream transmission.

preferable because the manufacturer can affect downstream incentives, whereas operators will have difficulty affecting upstream incentives.

As a final note, it is important to point out that the hypothesized percolation effect of residual liability on activity levels discussed in Corollary 3.5 will not undermine the manufacturer's incentives to produce safer robots, discussed in Corollary 3.2. Even when manufacturers are able to transfer the cost of their expected residual liability back to operators (e.g., charging "maintenance fees" equal to the residual liability associated with the activity level of the robot), their incentives to produce safer robots would remain in place. Newer, safer robots would in fact be cheaper and have lower "maintenance fees" and they would therefore be financially more attractive in the marketplace. The percolation effect of manufacturers' residual liability on operators' activity levels would thus be a contributing force to drive unsafe robots out of the market.

4.3 Manufacturers' Incentives to Prove the Negligence of Operators and Victims

Assigning the residual liability to the manufacturer incentivizes manufacturers to incorporate evidence technologies into their robots, increasing the operators' (and victims') incentives to adopt due care in their use and exposure to robots. To understand this additional effect of our rule, we should note that under a negligence regime, victims have the burden of proving the negligence of their injurers to obtain compensation. Compensation is a powerful incentive to produce evidence of negligence. However, under a rule of manufacturer residual liability, the victim's incentives to prove the injurer's negligence may be reduced. Under manufacturer residual liability, a victim can bring an action and obtain compensation, even if he/she fails to prove the negligence of the operator. In the absence of proof of the operator's negligence, the manufacturer is liable to compensate the (non-negligent) victim for the harm suffered.

However, fortunately the victim's reduced incentives to prove the operator's negligence do not undermine the operator's incentives to invest in optimal precautions. The proposed manufacturer residual liability rule may reduce the efforts of the victim to prove the negligence of the operator; however, the residually liable manufacturer will be incentivized to prove the negligence of the operator, and may have better access to information to do so. The victim's reduced litigation efforts will be (more than) fully offset by the manufacturer's efforts to prove the operator's negligence. Manufacturers will anticipate that victims will rely on the manufacturer's residual liability to obtain compensation and are therefore incentivized to invest resources to save evidence regarding the operator's activities in the hope of proving that the accident was caused by the operator's negligence. Under a manufacturer residual liability regime, manufacturers would likely incorporate evidence technologies to document the operator's use of the robot to limit their liability exposure in the event of accidents caused by the operator's negligence.

In the context of robot torts, evidence technology can be used to document the operator's use of the product. For example, black box recording technology, dash-cams, and telematic and location-tracking technology could be installed in self-driving cars (see Guerra and Parisi, 2020). Similarly, a decision log can be installed in a surgical robot to record all of the decisions that the robot made while performing an operation. This is an improvement over the typical human tort. Manufacturers of robots have a comparative advantage over victims in documenting and proving the operator's violation of safety standards.[49] Shifting litigation incentives from victims to manufacturers thus increases the probability that the negligent operators will face liability, thereby reinforcing their primary incentives to adopt optimal care.

When the manufacturer's residual liability operates under rules of contributory or comparative negligence (see Figure 1 and related explanation), manufacturers would also have incentives to monitor the care levels of victims, since under these rules finding negligence by victims would equally shield manufacturers from liability (Guerra and Parisi, 2020).

5 Final Remarks

Robot technologies are transforming the world, and technical progress is outpacing legal innovation. In this paper, we have provided a novel legal analysis of robot-generated accidents from an economic perspective. The development of general tort rules for accidents caused by robots will take

[49]Under a standard negligence rule, when the victim brings an action, he/she would generally bear the burden to prove the negligence of the operator to establish liability. In the case of robot torts, the proof could be more difficult than in ordinary cases. Imagine the use of a robot in a complex medical procedure or in an automated flight operation and the information required by the victim to prove the operator's negligence: this would make the victim's probability of satisfying the burden of proof lower, and the threat of negligence liability less effective in these situations.

time, but we hope that our analysis will help policymakers to think outside of the box when designing liability rules for robot-generated accidents. The main takeaway of our analysis is that a general, fault-based liability regime where operators and victims bear accident losses attributable to their negligent behavior and manufacturers are only held liable for non-negligent accidents may provide a second-best efficient set of incentives, accomplishing the four objectives of a liability regime: (1) incentivizing efficient care levels, (2) incentivizing efficient investments in developing safer robots, (3) incentivizing the adoption of newer and safer robots, and (4) incentivizing efficient activity levels.

The primary fault-based liability rules incentivize due care levels for operators and victims. The manufacturer's residual liability for non-negligent accidents incentivizes manufacturers to invest in safer robot technologies. The expected cost of the manufacturer's residual liability is incorporated in the price of the robot, driving unsafe robots out of the market and incentivizing the adoption of newer technology. The robots' ability to gather and record reliable data that can be used in court has crucial implications for the optimal allocation of residual liability. Robot manufacturers can install instruments that can effectively monitor the activity level and the unobservable precautions of the operator more effectively than victims or courts. Even though operators are not the direct bearers of residual liability, manufacturers will be able to transfer the cost of the residual liability to the robots' operators, incentivizing them to reduce their activity to socially optimal levels. Manufacturers would additionally have incentives to invest in technology to monitor and document the care levels of operators and victims in the hopes of proving their negligence, to avoid facing residual liability. This will reinforce the operators' and victims' incentives to act diligently.

However, the mere design of an applicable liability regime for robot technologies is not the only mechanism by which to incentivize further automation. There are also other means available, including regulation and mandatory adoption requirements, intellectual property rights, prizes, preferential tax treatments, or tax premiums. Insurance discounts for individuals adopting automated technologies can mitigate potentially high adoption costs. An optimal combination of these policy instruments may foster a widespread use of automated and safer technologies.

The focus of our paper has been how to incentivize manufacturers to constantly improve the design of their robots while keeping in place all of the other parties' primary incentives. As discussed in

Section 2.1, with some of the new programming techniques, the improvement of the robot is carried out by the robot itself, and robots can evolve beyond the design and foresight of their original manufacturers. With these technologies, we face what Matthias (2004) described as the "responsibility gap", whereby it is increasingly difficult to attribute the harmful behavior of "evolved" robots to the original manufacturer.

In this context, future work should consider an extension of our analysis, in which robots could become their own legal entity with financial assets attached to them, like a corporation.[50] Robots could be assigned some assets to satisfy future claims, and perhaps a small fraction of the revenues earned from the robot's operation could be automatically diverted to the robot's asset base, improving its solvency. Claims exceeding the robot's assets could then fall on the manufacturer under our residual liability rule.

At present, rules of manufacturer residual liability cannot be observed in practice and we can only offer theoretical conjectures on their operation. The effectiveness of these rules vis-a-vis other liability rules in creating optimal incentives may deserve some experimental and empirical investigation. Future scholarship should also analyze the effects of the allocation of residual liability to robot manufacturers on the development of possible pricing mechanisms to transmit incentives from manufacturers to operators, and how market relationships between operators and victims could align activity-level incentives also for victims. Moreover, in this case the effectiveness of pricing mechanisms to realign the parties' activity levels is a critical question that warrants experimental and empirical testing. Finally, future scholarship might consider how incentives of operators and manufacturers would change if robots were held responsible for accidents under a negligence standard, comparing its independently-performed activities and levels of safety to those of a reasonable person or "reasonable robot."

References

Abraham, K. S. and R. L. Rabin (2019). Automated vehicles and manufacturer responsibility for accidents: A new legal regime for a new era. *Virginia Law Review 105*, 127.

[50]This would not necessarily require the granting of legal personality to robots, as discussed in the European Parliament Report in Bertolini (2020). A simple bond or escrow requirement for robots likely to cause harm to third parties would suffice.

Allain, J. S. (2012). From Jeopardy to Jaundice: The medical liability implications of Dr. Watson and other artificial intelligence systems. *La. L. Rev. 73*, 1049.

Ben-Shahar, O. (2016). Should carmakers be liable when a self-driving car crashes. *Forbes*.

Bertolini, A. (2013). Robots as products: the case for a realistic analysis of robotic applications and liability rules. *Law, innovation and technology 5*(2), 214–247.

Bertolini, A. (2020). Artificial intelligence and civil liability. *Bruxelles: European Parliament – Committee on Legal Affairs 608*, 1–132.

Bertolini, A. and F. Episcopo (2021). The Expert Group's Report on Liability for Artificial Intelligence and Other Emerging Digital Technologies: A Critical Assessment. *European Journal of Risk Regulation*, 1–16.

Bertolini, A. and M. Riccaboni (2020). Grounding the case for a european approach to the regulation of automated driving: The technology-selection effect of liability rules. *European Journal of Law and Economics*, 1–42.

Burridge, N. (2017). Artificial intelligence gets a seat in the boardroom. *Nikkei Asian Review 10*.

Calo, M. (2011). Open robotics. *Maryland Law Review 42*, 101–142.

Calo, R. (2015). Robotics and the lessons of cyberlaw. *California Law Review 103*, 513.

Calo, R., A. M. Froomkin, and I. Kerr (2016). *Robot Law*. Edward Elgar Publishing.

Carbonara, E., A. Guerra, and F. Parisi (2016). Sharing residual liability: The cheapest cost avoider revisited. *The Journal of Legal Studies 45*(1), 173–201.

Carroll, K. (2021). Smart cars are getting smarter: Legal personhood for self-driving vehicles. *Working Paper Seton Hall University eRepository*.

Casey, B. (2019). Robot ipsa loquitur. *Georgetown Law Journal 108*, 225.

Casey, B. and M. A. Lemley (2019). You might be a robot. *Cornell Law Review 105*, 287.

Chopra, S. and L. F. White (2011). *A legal theory for autonomous artificial agents*. University of Michigan Press.

Colonna, K. (2012). Autonomous cars and tort liability: Why the market will "drive" autonomous cars out of the marketplace. *Journal of Law, Technology, & the Internet 4*(1), 81.

Cooling, J. E. and P. V. Herbers (1983). Considerations in autopilot litigation. *Journal of Air Law and Commerce 48*, 693.

Cooter, R. and A. Porat (2013). Lapses and substitution. In R. Cooter and A. Porat (Eds.), *Getting Incentives Right: Torts, Contracts and Restitution*, pp. 61–73. Princeton: Princeton University Press.

Crane, D. A., K. D. Logue, and B. C. Pilz (2017). A survey of legal issues arising from the deployment of autonomous and connected vehicles. *Michigan Telecommunications and Technology Law Review 23*, 191.

Dari-Mattiacci, G. and L. A. Franzoni (2014). Innovative negligence rules. *American Law and Economics Review 16*(2), 333–365.

Dari-Mattiacci, G., B. Lovat, and F. Parisi (2014). Loss-sharing between nonnegligent parties. *Journal of Institutional and Theoretical Economics: JITE*, 571–598.

Duda, R. O. and E. H. Shortliffe (1983). Expert systems research. *Science 220*(4594), 261–268.

Eish, M. and T. Hwang (2015). Praise the Machine! Punish the Human! *Comparative Studies in International Systems, Working Paper* (1).

Epstein, R. A. (2021). Liability rules for autonomous vehicles. *Journal of Entrepreneurship and Public Policy*.

Esteve, J. S. and C. B. Domènech (2017). Rights and science in the drone era: actual challenges in the civil use of drone technology. *Rights and Science: R&S*, 117–133.

European Parliament (2017). European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103-INL). *Estrasburgo: Parlamento Europeu*.

Evas, T., C. Rohr, F. Dunkerley, and D. Howarth (2018). A Common EU Approach to Liability Rules and Insurance for Connected and Autonomous Vehicles: European Added Value Assessment Accompanying the European Parliament's Legislative Own-Initiative Report. *EU Publications*.

Fagnant, D. J. and K. Kockelman (2015). Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice 77*, 167–181.

Fox, J., J. North, and J. Dean (2019). Ai in the boardroom: Could robots soon be running companies? *Governance Directions 71*(10), 559–564.

Garza, A. P. (2011). Look Ma, No Hands: Wrinkles and Wrecks in the Age of Autonomous Vehicles. *New England Law Review*.

Gilo, D. and E. Guttel (2009). Negligence and insufficient activity: The missing paradigm in torts. *Michigan Law Review 108*, 277.

Giuffrida, I. (2019). Liability for ai decision-making: some legal and ethical considerations. *Fordham Law Review 88*, 439–442.

Giuffrida, I., F. Lederer, and N. Vermeys (2017). A legal perspective on the trials and tribulations of ai: How artificial intelligence, the internet of things, smart contracts, and other technologies will affect the law. *Case Western Reserve Law Review 68*, 747.

Grady, M. (1987). Why Are People Negligent Technology, Nondurable Precautions, and the Medical Malpractice Explosion. *Northwestern University Law Review 1*.

Grady, M. F. (2009). Unavoidable Accident. *Review of Law & Economics 5*(1).

Graham, K. (2012). Of Frightened Horses and Autonomous Vehicles: Tort Law and its Assimilation of Innovations. *Santa Clara Law Review 52*(4).

Guerra, A. and F. Parisi (2020). Investing in private evidence. *Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3596029*.

Hay, B. and K. E. Spier (2005). Manufacturer liability for harms caused by consumers to others. *American Economic Review 95*(5), 1700–1711.

Hubbard, F. P. (2014). Sophisticated robots: Balancing liability, regulation, and innovation. *Florida Law Review 66*, 1803.

Imai, T. (2019). Legal regulation of autonomous driving technology: Current conditions and issues in Japan. *IATSS research 43*(4), 263–267.

Jones, C. (2018). The Robot Koseki: A Japanese Law Model for Regulating Autonomous Machines. *J. Bus. & Tech. L. 14*, 403.

Jones, T. (2017). International commercial drone regulation and drone delivery services. Technical report, RAND.

Kamensky, S. (2020). Artificial intelligence and technology in health care: Overview and possible legal implications. *DePaul Journal of Health Care Law 21*(3), 3.

Kassahun, Y., B. Yu, A. T. Tibebu, D. Stoyanov, S. Giannarou, J. H. Metzen, and E. Vander Poorten (2016). Surgical robotics beyond enhanced dexterity instrumentation: a survey of machine learning techniques and their role in intelligent and autonomous surgical actions. *International Journal of Computer Assisted Radiology and Surgery 11*(4), 553–568.

Kop, M. (2019). AI & intellectual property: Towards an articulated public domain. *Texas Intellectual Property Law Journal 28*, 297.

Kovač, M. (2020). *Judgement-Proof Robots and Artificial Intelligence: A Comparative Law and Economics Approach*. Palgrave Macmillan Springer Nature.

Księżak, P. and S. Wojtczak (2020). AI versus robot: in search of a domain for the new European civil law. *Law, Innovation and Technology 12*(2), 297–317.

Lanfranco, A. R., A. E. Castellanos, J. P. Desai, and W. C. Meyers (2004). Robotic surgery: A current perspective. *Annals of surgery 239*(1), 14.

Leis, M. J. (2006). *Robots-our future partners?!: A sociologist's view from a German and Japanese perspective*. Tectum-Verlag.

Lemley, M. A. and B. Casey (2019). Remedies for robots. *The University of Chicago Law Review 86*(5), 1311–1396.

Leveen, S. A. (1983). Cockpit controversy: The social context of automation in modern airliners. *Ph.D. Dissertation. Cornell University, Department of Science and Technology Studies*.

Luppi, B., F. Parisi, and D. Pi (2016). Double-edged torts. *International Review of Law and Economics 46*, 43–48.

MacDorman, K. F., S. K. Vasudevan, and C.-C. Ho (2009). Does Japan really have robot mania? Comparing attitudes by implicit and explicit measures. *AI & society 23*(4), 485–510.

Marchant, G. E. and R. A. Lindor (2012). The Coming Collision Between Autonomous Vehicles and the Liability System. *Santa Clara Law Review 52*(4).

Marchant, G. E. and L. M. Tournas (2019). AI Health Care Liability: From Research Trials to Court Trials. *Journal of Health & Sciences Law 12*, 23–41.

Matsuzaki, H. and G. Lindemann (2016). The autonomy-safety-paradox of service robotics in Europe and Japan: A comparative analysis. *AI & Society 31*(4), 501–517.

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology 6*(3), 175–183.

Mele, D. M. (2013). The Quasi-Autonomous Car as an Assistive Device for Blind Drivers: Overcoming Liability and Regulatory Barriers. *Syracuse Science & Technology Law Reporter*, 26–64.

Miceli, T. J. (1997). *Economics of the law: Torts, contracts, property, litigation*. Oxford: Oxford University Press.

Miceli, T. J. (2017). *An Economic Model of Accidents: The Model of Precaution*, pp. 18–37. Stanford University Press.

Michalski, R. (2018). How to sue a robot. *Utah Law Review*, 1021.

Mingtsung, C. and Q. Wei (2020). Research on infringement of artificial intelligence medical robot. In *2020 4th International Seminar on Education, Management and Social Sciences*, pp. 497–500. Atlantis Press.

Möslein, F. (2018). Robots in the boardroom: artificial intelligence and corporate law. In *Research handbook on the law of artificial intelligence*. Edward Elgar Publishing.

Mot, J. D. and B. Depoorter (2011). Technology and torts: Memory costs, nondurable precautions and interference effects. *International Review of Law and Economics*.

Mulligan, C. (2017). Revenge against robots. *SCL Rev. 69*, 579.

National Highway Traffic Safety Administration (2013). Preliminary statement of policy concerning automated vehicles.

National Highway Traffic Safety Administration (2016). Federal automated vehicles policy. *Official Policy*.

Nussim, J. and A. D. Tabbach (2009). A revised model of unilateral accidents. *International Review of Law and Economics 29*(2), 169–177.

Polinsky, A. M. and Y.-K. Che (1991). Decoupling liability: optimal incentives for care and litigation. Technical report, National Bureau of Economic Research.

Rachum-Twaig, O. (2020). Whose robot is it anyway?: Liability for artificial-intelligence-based robots. *U. Ill. L. Rev.*, 1141.

Roe, M. (2019). Who's driving that car: An analysis of regulatory and potential liability frameworks for driverless cars. *BCL Rev. 60*, 317.

Šabanović, S. (2014). Inventing Japan's 'robotics culture': The repeated assembly of science, technology, and culture in social robotics. *Social Studies of Science 44*(3), 342–367.

Schellekens, M. (2015). Self-driving cars and the chilling effect of liability law. *Computer Law & Security Review 31*(4), 506–517.

Shavell, S. (1980). Strict liability versus negligence. *Journal of Legal Studies 9*(1), 1–25.

Shavell, S. (1987). *Economic Analysis of Accident Law*. Cambridge, Massachusetts: Harvard University Press.

Shavell, S. (2008). On optimal legal change, past behavior, and grandfathering. *The Journal of Legal Studies 37*(1), 37–85.

Shavell, S. (2020). On the redesign of accident liability for the world of autonomous vehicles. *The Journal of Legal Studies 49*(2), 243–285.

Sheets, K. D. (2018). The japanese impact on global drone policy and law: Why a laggard united states and other nations should look to japan in the context of drone usage. *Indiana Journal of Global Legal Studies 25*, 513.

Smith, B. W. (2014). Automated Vehicles Are Probably Legal in the United States. *Texas A&M Law Review 1*(3), 411–521.

Stöcker, C., R. Bennett, F. Nex, M. Gerke, and J. Zevenbergen (2017). Review of the current state of UAV regulations. *Remote Sensing 9*(5), 459.

Talley, E. (2019). Automatorts: How Should Accident Law Adapt to Autonomous Vehicles? Lessons from Law and Economics. *Draft available at: https://www.hoover.org/sites/default/files/ip2-19002-paper.pdf.*.

Weston, S. B. (1963). Contributory negligence in product liability. *Cleveland-Marshall Law Review 12*, 424.

World Economic Forum (2015). Deep shift: technology tipping points and societal impact.

Zolfagharifard, E. (2014). Would you take orders from a robot? An artificial intelligence becomes the world's first company director. *Daily Mail*.

Appendix A

Proof of Proposition 3.1. Proving this proposition requires us to show that (1) the injurer and the victim have incentives to invest in at least due care $\forall\theta \in [0,1]$; and (2) the manufacturer has socially optimal incentives to invest in R&D and production of safer robots, r, $\forall\theta \in [0,1]$.
To prove point (1), we need to show that neither party has incentives to deviate from the due-care equilibrium by investing in sub-optimal care. In equilibrium, the injurer is not able to increase his/her pay-off by undertaking less than due care ($x < x^{**}$) if the following condition holds:

$$V_O(w^*) - w^*x^{**} > V_O(w) - \frac{wz^{**}}{r^{**}} p(x, y^{**})L - wx \tag{A.1}$$

with $x < x^{**}$, $w^* = w(x^{**}, y^{**})$, $w = w(x, y^{**})$. Given that x^{**} maximizes the social welfare (3.1), we must have that, for any $x < x^{**}$,

$$\begin{aligned} V_O(w^{**}) + V_V(z^{**}) - \tfrac{w^{**}z^{**}}{r^{**}} p(x^{**}, y^{**})L - w^{**}x^{**} - z^{**}y^{**} - r^{**} > \\ V_O(w) + V_V(z^{**}) - \tfrac{wz^{**}}{r^{**}} p(x, y^{**})L - wx - z^{**}y^{**} - r^{**} \end{aligned} \tag{A.2}$$

Rearranging:

$$V_O(w^{**}) - \frac{w^{**}z^{**}}{r^{**}} p(x^{**}, y^{**})L - w^{**}x^{**} > V_O(w) - \frac{wz^{**}}{r^{**}} p(x, y^{**})L - wx \tag{A.3}$$

Given that $w^* > w^{**}$ (see proof of Proposition 3.4), $V_O(w^*) > V_O(w^{**})$. Thus, $V_I(w^*) - w^*x^{**} > V_O(w^{**}) - \frac{w^{**}z^{**}}{r^{**}} p(x^{**}, y^{**})L - w^{**}x^{**}$. Given (A.3), we obtain $V_I(w^*) - w^*x^{**} > V_O(w) - \frac{wz^{**}}{r^{**}} p(x, y^{**})L - wx$, with $x < x^{**}$ and $w = w(x, y^{**})$, which proves that undertaking less than due care is not an equilibrium. This is true $\forall\theta \in [0,1]$, and a similar proof holds for the victim.
Regarding point (2), in equilibrium, the manufacturer minimizes $r + \frac{w^*z^*}{r} p(x^{**}, y^{**})L$. The minimization problem yields $r^* : \frac{wz}{r^2} pL = 1$, which is equal to the socially optimal level as defined in Equation (3.6). Intuitively, since w, z cannot be included in the standard of due care, each party will have full incentives on these variables only if they are the residual bearer of the loss. Thus, $z^* = z^{**}$ only if residual liability is entirely borne by the victim and $w^* = w^{**}$ only if residual liability is entirely borne by the robot operator. Neither of these allocations of residual liability is compatible with the allocation of residual liability needed to incentivize optimal $r^* = r^{**}$, as shown in point (3) below.

□

Proof of Proposition 3.4. Proving this proposition requires us to show that the injurer and the victim may have incentives to undertake excessive activity levels. In equilibrium ($x^* = x^{**}$ and $y^* = y^{**}$), the injurer maximizes $V_I(w^*) - w^* x^{**}$, with $w^* = w(x^{**}, y^{**})$. Since $V_I' > 0$ and $V_I'' \leq 0$, and there are no liability costs, the injurer will have incentives to over-invest in activity levels, $w^* > w^{**}$. A similar reasoning applies for victims.

Consistent with Shavell's (1980) activity-level theorem, the operator and victim will have full incentives to choose socially optimal values of w and z only if they are the residual bearers of the loss. Thus, in a regime of manufacturer residual liability in which the robot's manufacturer is the sole residual bearer, it follows that $z^* > z^{**}$ and $w^* > w^{**}$. This misalignment of activity-level incentives is necessary to incentivize optimal $r^* = r^{**}$, as shown in point (2) above. □