


RESEARCH ARTICLE

Liability for robots II: an economic analysis

Alice Guerra^{1,2*} , Francesco Parisi^{2,3} and Daniel Pi⁴

¹Department of Economics, University of Bologna, via Angherà 22, 47921 Rimini, Italy, ²Department of Economics, University of Bologna, Bologna, Italy, ³School of Law, University of Minnesota, Minneapolis, Minnesota, USA and ⁴School of Law, University of Maine, Portland, Maine, USA

*Corresponding author. Email: alice.guerra3@unibo.it

(Received 20 August 2021; revised 1 November 2021; accepted 2 November 2021; first published online 29 November 2021)

Abstract

This is the second of two companion papers that discuss accidents caused by robots. In the first paper (Guerra *et al.*, 2021), we presented the novel problems posed by robot accidents, and assessed the related legal approaches and institutional opportunities. In this paper, we build on the previous analysis to consider a novel liability regime, which we refer to as ‘manufacturer residual liability’ rule. This makes operators and victims liable for accidents due to their negligence – hence, incentivizing them to act diligently; and makes manufacturers residually liable for non-negligent accidents – hence, incentivizing them to make optimal investments in R&D for robots’ safety. In turn, this rule will bring down the price of safer robots, driving unsafe technology out of the market. Thanks to the percolation effect of residual liability, operators will also be incentivized to adopt optimal activity levels in robots’ usage.

Key words: Automated technology; liability; manufacturer residual liability; negligence

JEL codes: K13; K32

1. Introduction

Machine decision-making reduces the impact of human error but does not entirely eliminate the risk of accidents, creating fresh dangers in the form of machine malfunctions and design limitations (Bertolini, 2015; Bertolini *et al.*, 2016; De Chiara *et al.*, 2021; European Commission, 2020). In the companion paper (Guerra *et al.*, 2021), we have detailed the novel, legal problems posed by robot accidents, and extensively discussed the reasons why existing tort rules – standard negligence law and products liability – do not adequately address the multi-dimensional incentive issues that arise in this novel setting. We have shown that, to date, there exists no general formulation of liability in case of robot accidents, and the proposed solutions differ across jurisdictions. The still open question concerns how legal incentives – so carefully calibrated to induce *human* actors to exercise precautionary care – need to be reshaped in robot accidents (Bertolini, 2013, 2020; Bertolini and Riccaboni, 2020; Casey and Lemley, 2019; De Chiara *et al.*, 2021; Epstein, 2021; Kovač, 2020; Lemley and Casey, 2019; Shavell, 2020; Talley, 2019).

Here, we build on these legal challenges and propose – through a formal economic model – a novel liability regime, which we refer to as ‘manufacturer residual liability’ (‘MRL’, hereinafter). This regime applies to robots operated with human intervention, and shifts liability to manufacturers provided that operators and third-party victims have invested in due care.¹ This rule is the three-way analog of a

¹Our MRL rule 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 the case of the insolvency of product users who caused harm

© The Author(s), 2021. Published by Cambridge University Press on behalf of Millennium Economics Ltd.. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

manufacturers' strict liability rule, coupled with a defense of contributory negligence, although here the negligence defense applies to both robot operators and victims. The paper shows that an MRL rule could be an effective liability regime for robot torts. Specifically, MRL provides a second-best efficient set of incentives, accomplishing four objectives: incentivizing (1) efficient levels of human care by operators and victims, (2) efficient activity levels in the use of robots; (3) efficient R&D investments for the development of safer robots; and (4) adoption of safer robots in the marketplace. We further analyze the effects of an MRL rule on the parties' incentives to gather and produce evidence for the adjudication of robot torts.

This paper comprises five sections. Section 2 reviews the literature. Section 3 presents a formal model of MRL rules, whose effects are discussed in section 4. Section 5 concludes the paper.

2. Related literature

Robot technologies are increasingly used in a variety of different settings, and technical progress is in many ways outpacing legal innovation. In the companion paper (Guerra *et al.*, 2021), we survey some representative implementations to assess how legal rules and institutions have responded to the presence of robot actors to date. Presently, there exists no general formulation of liability in case of accidents caused by robots, and the proposed solutions differ across jurisdictions.

In the economic analysis of tort law, there have been several forays into analyzing robot torts, with special attention to self-driving cars (e.g. De Chiara *et al.*, 2021; Fagnant and Kockelman, 2015; Roe, 2019; Shavell, 2020). Most robot-generated accidents have been analyzed under a framework of products liability, where the victim is the product user or a third party, and liability for the harm is assigned to manufacturers (e.g. Abraham and Rabin, 2019; Ben-Shahar, 2016; Crane *et al.*, 2017; Evas *et al.*, 2018). However, as discussed in Guerra *et al.* (2021), approaching the problem from a products liability perspective assumes many of the answers, but the conceptual problem is much more general. The products liability focus on product malfunctions which are the result of design or manufacturing defects, and could have been avoided by R&D investments aimed at increasing the safety of current technology. Our MRL rule – which is not a rule of products liability – encompasses a broader range of accident situations (see section 3.1), by holding manufacturers strictly liable when operators and victims are not negligent regardless of design or manufacturing defects.

Three existing papers have approached the tort problem at a greater level of generality: Lemley and Casey (2019); Talley (2019); and Shavell (2020). In particular, 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 MRL rule is simpler and more practicable to optimally align incentives for all three parties in robot torts.

Shavell (2020) proposes a rule of strict liability for automated vehicles and manufacturers with damages payable to the state. This is equivalent to a decoupling rule (Polinsky and Che, 1991). Our rule substantially differs from Shavell's, in that manufacturers face residual strict liability, and damages are paid to faultless victims rather than the state. The advantage is two-fold. First, our rule leads to payment of damages to the actual victims of the accident rather than the state, hence serving 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

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 MRL rule applies to the class of negligence rules, under which liability does depend upon the precautions taken by operators and victims (see section 3.1). Further in contrast to Hay and Spier (2005), the efficiency of our rule does not depend upon the operator's financial assets. It rather holds for the broader class of harms caused by non-negligent operators. Given the different premises and scopes of the two rules, the incentives that they generate are also different, as discussed throughout the paper. On product-related accidents with third-party victims, see also Polinsky (1980).

the liability rule by engaging in litigation. As is well known (Polinsky and Che, 1991), these incentives may be diluted under decoupling rules, because in practice the injured party has no incentive to sue.

Although the specific question of liability for robots has only been considered by a few contributions, 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 (e.g. Grady, 1987, 2009; Mot and Depoorter, 2011). Tort scholars have also studied incentives to upgrade to newer, safer technologies.² Among others, Dari-Mattiacci and Franzoni (2014) analyze 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.

3. A theory of manufacturer residual liability

3.1 Terminology and scope

The scenarios we analyze to delimit the scope of our inquiry arise from the interaction of three parties: an operator, a victim, and a manufacturer. We use the term ‘operator’ to refer to the human who utilizes the robot to carry out an activity. Our analysis assumes the existence of a human operator who has some form of control over the robot – control which can be exercised in a negligent manner. The control can range from situations where the operator directs the activity of the robot and can override its decisions, to situations where the role of the operators is only marginal, such as periodical maintenance, monitoring warning messages sent by the robot, and avoidance of use robot in hazardous environmental conditions. We thus exclude ‘humans-out-of-the-loop’ robots that are completely capable of self-determination.³ In the extreme case of fully-automated robots, operators would have no control over the robot and their incentives would no longer be a relevant focus in the analysis.⁴

We use the term ‘victim’ to refer to a third party who has suffered harm caused by a robot’s decisions. We are interested in general circumstances where the operator and victim are distinct individuals. The specific case where the operator is also the victim reduces to an ordinary case of products liability. We use the term ‘manufacturer’ to refer to the entity responsible for the development, production, and sale of a robot. There may be several parties embedded within our ‘manufacturer’, including the robot programmer. Our analysis seeks to answer only whether and how liability should be assigned to a *monolithic* manufacturer. The internal allocation of liability among the multiple parties clustered in the ‘manufacturer’ as a firm fall 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 contrary, if an automobile crashes into a pedestrian 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

²The 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]).

³Although there exist different types of robots, and the level of autonomy of each robot type exists on a spectrum (Bertolini *et al.*, 2016), their immediate legal consequences are binary (Casey, 2019): the actions that led to the accidents either involved tasks where robots operated with human intervention (‘humans-in-the-loop’ activities), or tasks that neither involved nor needed human intervention (‘humans-out-of-the-loop’ activities).

⁴In that case, the operator’s role would vanish and the only parties remaining would be the robot’s manufacturer and the robot’s victim. In this limiting case, our MRL rule would reduce to a rule of manufacturer strict liability coupled with a defense of contributory applied against third-party victims. And if victims adopted fully automated robots as well (e.g. in a world with only self-driving cars), their care choices would also become irrelevant – reducing our analysis to a conventional manufacturer strict liability regime. Instead, in our analysis, operators have an active role that can affect the risk created by the robot’s activity. We thank Gerhard Wagner and an anonymous reviewer for this comment.

in other (non-decision-making) components of the machine cause harm, the accident happens to involve a robot, but falls outside the scope of our analysis. Here, we are interested in cases where the robot's choices cause harm. Furthermore, what marks the boundary between an operator's (or victim's) negligence liability and a manufacturer's residual liability in our proposed rule is that the result of an accident was not caused by the negligence of the robot's operator (or victim).

We shall also summarize those features of robot torts that are distinguishable from human torts – which we have extensively explained in Guerra *et al.* (2021). 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: a robot may be partially autonomous, sharing decision-making duties with a human operator; or fully autonomous but the operator still decides its objectives and may be allowed 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. 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 be subsequently 'forgotten'.

Under our proposed liability rule, an operator's negligence could materialize in a variety of ways. Examples include the misuse of the robot (e.g. using a medical robot on the wrong patient) or use outside its designed range of applications (e.g. use of vessels in prohibitive weather conditions); software or hardware alterations; lack of required maintenance; and negligent control (e.g. ignoring alerts sent by the robot to the operator). Operator's primary liability would arise when such negligence can be identified as the main cause of the harm to a victim.

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.⁵ Our rule would additionally allow us to tackle the challenges that emerge in the regulation of robots' 'design limitations', i.e. accidents that occur when the robot encounters a new, unforeseen circumstance that causes it to behave in an undesired manner. Most accidents caused by 'design limitations' could be avoided with greater investments in R&D and/or safety updates. As shown in section 3.2, our MRL rule outperforms rules of products liability for product malfunction in this dimension, by incentivizing manufacturers to constantly improve the design of their robots, while keeping in place all of the other parties' primary incentives.

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. 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 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), MRL rules 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.⁶ Under a contributory negligence rule, a

⁵For additional examples on the type of risks that robots can generate, see Bertolini *et al.* (2016).

⁶Simple 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).

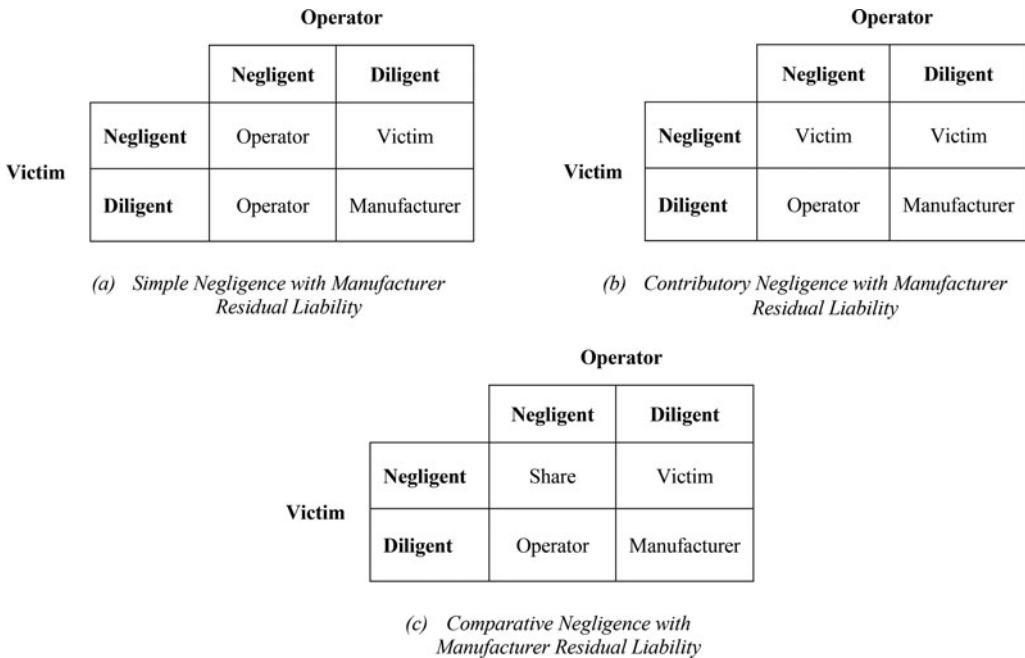


Figure 1. MRL rules.

negligent victim would be completely barred from obtaining compensation in a negligence case.⁷ Under a comparative negligence rule, when both operators and victims are at least somewhat negligent, the accident loss would be divided between them as a percentage based on their relative degree of negligence.⁸ As 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, we jointly discuss the incentive effects of all three MRL rules presented above.

3.2 A basic model

We denote the human operator by O , the third-party victim by V , and the robot manufacturer by M . Let x and w represent the operator's care and activity levels, respectively⁹; and y and z the victim's care and activity levels, respectively. The value of the activity is denoted by $V_O(w)$ for the operator, with $V'_O > 0$, $V''_O \leq 0$; and by $V_V(z)$ for the victim, with $V'_V > 0$, $V''_V \leq 0$. We do not consider the availability of robot technology for the victim.¹⁰ Let r represent the manufacturer's investments to produce the robot. Production costs include R&D investments for the development and improvements in the quality and safety of the robot. The accident loss is represented by $L > 0$.

⁷Contributory 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.

⁸Several jurisdictions in the USA 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.

⁹The operator's activity level is intended as usage per unit purchased. See section 4.2.

¹⁰As the probability of symmetrical use arises (where both the injurer and victim 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 directly follows from Luppi *et al.* (2016).

At this point, there are several ways to model robot torts. The most general approach would be to define the social cost function by $L(x, y, w, z, r) + wx + zy + r$, where $L(\cdot)$ denotes the expected accident loss. Following the conventional setup (Shavell, 1980; 1987), $L(\cdot)$ is decreasing and convex in care levels, and increasing and weakly convex in activity levels. Formally, $\partial L/\partial x < 0$, $\partial L/\partial y < 0$, $\partial^2 L/\partial x^2 > 0$, $\partial^2 L/\partial y^2 > 0$, $\partial L/\partial w > 0$, $\partial L/\partial z > 0$, $\partial^2 L/\partial w^2 \geq 0$, $\partial^2 L/\partial z^2 \geq 0$.¹¹ We assume that greater R&D investments reduce both the expected accident loss with diminishing marginal effectiveness – i.e. $\partial L/\partial r < 0$, $\partial^2 L/\partial r^2 \geq 0$ – and the effectiveness of human precautionary care – i.e. $\partial^2 L/\partial x\partial r \geq 0$, $\partial^2 L/\partial y\partial r \geq 0$.

In the conventional setup, the activity levels generally enter the social cost function in the multiplicative form, i.e. $wzp(x, y)L + wx + zy$, where $p(x, y)$ represents the probability of an accident which is decreasing and convex in both x and y (Landes and Posner, 1987; Miceli, 1997; Shavell, 1980, 1987). This captures the fact that both the operator and the victim need to carry out the activity for the accident to occur. For the ease of comparison and tractability, without loss of generality we express the social cost function as $(wz/r)p(x, y)L + wx + zy + r$.¹² This representation 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 term r at the denominator captures the assumption that increasing investments in researching robot technology reduces the expected accident loss; and (3) investments in r reduce the effectiveness of human precautionary care, that is, automation does not supplement but rather replaces human precautionary care.

3.3 Social optimization problem

The social optimization problem is:

$$\max_{x,y,w,z,r} S = V_O(w) + V_V(z) - \frac{wz}{r}p(x, y)L - wx - zy - r. \tag{1}$$

The operator’s and victim’s efficient care levels are:

$$x^{**}: -\frac{wz}{r} \frac{\partial p}{\partial x} L = w \tag{2}$$

$$y^{**}: -\frac{wz}{r} \frac{\partial p}{\partial y} L = z. \tag{3}$$

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

The parties’ efficient activity levels are:

$$w^{**}: V'_O = \frac{z}{r}pL + x \tag{4}$$

$$z^{**}: V'_V = \frac{w}{r}pL + y. \tag{5}$$

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

The efficient research investment is:

$$r^{**}: \frac{wz}{r^2} pL = 1. \tag{6}$$

¹¹For similar formulations, see Singh (2006), Dari-Mattiacci *et al.* (2014), and Carbonara *et al.* (2016).

¹²This formulation does not affect our qualitative results, which still hold when using more general expected accident loss function, e.g. $L(x, y, w, z, r)$, or $wzL(x, y, r)$. We thank an anonymous referee for this comment.

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 analyze the private welfare functions.

3.4 Private optimization problem

We earlier identified three species of MRL rules, including simple negligence, contributory negligence, and comparative negligence (Figure 1). For ease of exposition, we introduce θ to represent the operator’s share of liability when both the operator and the victim are negligent. Under simple negligence with MRL, $\theta = 1$; under contributory negligence with MRL, $\theta = 0$; under comparative negligence with MRL, $0 < \theta < 1$. The private optimization problems of the operator, victim, and manufacturer, respectively, are:

$$\max_{x,w} U_O = \begin{cases} V_O(w) - wx & \text{if } \{x \geq x^{**} \wedge y < y^{**}\} \vee \{x \geq x^{**} \wedge y \geq y^{**}\} \\ V_O(w) - \theta \frac{wz}{r} pL - wx & \text{if } x < x^{**} \wedge y < y^{**} \\ V_O(w) - \frac{wz}{r} pL - wx & \text{if } x < x^{**} \wedge y \geq y^{**} \end{cases} \tag{7}$$

$$\max_{y,z} U_V = \begin{cases} V_V(z) - zy & \text{if } \{x < x^{**} \wedge y \geq y^{**}\} \vee \{x \geq x^{**} \wedge y \geq y^{**}\} \\ V_V(z) - (1 - \theta) \frac{wz}{r} pL - zy & \text{if } x < x^{**} \wedge y < y^{**} \\ V_V(z) - \frac{wz}{r} pL - zy & \text{if } x \geq x^{**} \wedge y < y^{**} \end{cases} \tag{8}$$

$$\min_r U_M = \begin{cases} r & \text{if } x < x^{**} \vee y < y^{**} \\ r + \frac{wz}{r} pL & \text{if } x \geq x^{**} \wedge y \geq y^{**}. \end{cases} \tag{9}$$

By solving the private optimization problems, under an MRL rule combined with any of the three negligence-based regimes (i.e. simple negligence, contributory negligence, and comparative negligence, hereafter jointly referred to as MRL rules), we obtain the following results¹³:

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

Proof. See the Appendix. □

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 more extensively discussed in section 4.1.1, under MRL rules, 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: 1701). Instead, this simply follows from considering that the manufacturer’s R&D investments are durable, i.e. once the safer algorithm has been developed,

¹³Consistent with convention, these optimality results are in fact second-best results, inasmuch there are second-order effects of excessive activity levels on care expenditures (Carbonara *et al.*, 2016; Dari-Mattiacci *et al.*, 2014).

no additional investment is required to maintain that safety level. Specifically, under MRL rules, manufacturers have incentives to minimize their total production costs also by including in r the cost of robot updates and maintenance. Their total expected investment in r – which includes their expected investment in post-sale safety research – will be up to the point where $(wz/r^2)pL = 1$, which is the socially optimal level of r as defined in equation (6). By adopting these socially optimal levels of R&D investments, 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. 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, $(wz/r)pL$. In a perfectly competitive market, the manufacturer maximizes profits by setting the price equal to the marginal cost, i.e. $(wzpL/r^2)$. 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 more extensively discussed in section 4.1.2, MRL rules 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. When using a very safe robot, the socially optimal levels of human care as defined in equations (2) and (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. \square

From equations (7) and (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 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 following:

$$\max_w U_O = V_O(w) - w(\tau + x^*) \quad (10)$$

where τ is the fee charged by the manufacturer to the operator per activity level (e.g. fee per mileage), and $x^* = x^{**}$ is defined as in equation (2). The manufacturer will set τ equal to the cost of residual liability per operator's activity level, i.e. $\tau = (z/r)pL$. The operator's privately optimal activity level, \bar{w}^* , becomes:

$$\bar{w}^*: V'_O = \frac{z}{r}pL + x^* \quad (11)$$

It follows that \bar{w}^* converges to the socially optimal level w^{**} as defined in equation (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.

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 is two-fold. First, when manufacturers face residual liability, they 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. 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.¹⁴ 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 about the costs and benefits of technological safety. Second, we should incentivize operators to adopt the safer robot technologies. Hence, a desirable liability rule 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 MRL rules 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.¹⁵ However, as discussed in sections 4.1.1 and 4.1.2, MRL rules can overcome this difficulty, creating incentives for manufacturers to produce safer and cheaper robots, and for operators to purchase them.

¹⁴However, see Dari-Mattiacci and Franzoni (2014), suggesting the possibility that negligence standards could be adjusted upward or downward when adoption costs are not verifiable in court, depending on whether the adopted technology reduces or increases expected harm.

¹⁵Most 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.

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: (1) that consumers are willing to pay a premium for safer products, and (2) that product safety is perfectly observable to consumers at the time when they make their purchasing decisions (Hay and Spier, 2005; Polinsky, 1980). However, under MRL rules, neither of these assumptions is likely to hold.

First, in the three-party scenario we consider, operators are only interested in avoiding liability, which they can do by adopting due care in using 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 (Guerra *et al.*, 2021).

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 non-negligent harm caused by the robot imposes an expected cost of liability on the manufacturer equal to $(wz/r)pL$. 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 – manufacturers 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 usage by operators, hence its potential future liability. Corollary 3.2 shows that under MRL rules, manufacturers make their production decisions based on the total cost that they face $K = r + (wz/r)pL$, rather than looking at the bare development cost, r , that they would face in the absence of residual liability. This induces manufacturers to invest in safety research until the last dollar spent reduces expected injury costs by one dollar, which is the socially optimal level of r as defined in equation (6). The resulting automated technology is therefore optimally safe.

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. Hence, under all three MRL rules, 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 in Corollary 3.3, MRL rules 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 MRL rules, manufacturers of older robots face higher levels of expected liability, and will therefore be strongly incentivized to find ways to upgrade or replace their obsolete robots. This may be accomplished, e.g. by offering maintenance plans or providing 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 replace obsolete robots and encourage operators to adopt the safest ones.

Even if manufacturers are incentivized to 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 most obvious solution would be to construe the operators' use of obsolete robots as *per se* negligence (Dari-Mattiacci and Franzoni, 2014). 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.

Overall, our proposed MRL rules achieve 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. Since manufacturers are in a better position than operators to control the safety of robots, they should be assigned sole residual liability.

4.2 Percolation effects on activity levels

Regardless of which species of MRL regime is chosen, negligence rules cannot incentivize non-verifiable precautions because non-verifiable precautions are – by definition – undetectable by courts in determining negligence (Shavell, 1980). In traditional accident cases involving an injurer and a victim, the creation of incentives for 'non-verifiable precautions' is accomplished through the allocation of residual liability.¹⁶ In our three-party scenario with operators, victims, and manufacturers, different parties have control over different aspects of non-verifiable precautions. Operators control their usage of the robot, manufacturers control R&D investments, and victims control their own activity levels.

By allocating the residual liability to manufacturers, our rule only creates direct incentives for optimal 'non-verifiable' R&D investments, leading to safer robots. 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.¹⁷

¹⁶Non-verifiable precautions include the reduction in the frequency of an activity, looking in the rear-view mirror while driving a car, investments in research for making activities safer. In the standard tort model, residual liability should be imposed on the party whose non-verifiable precautions most effectively reduce the cost of accidents (Carbonara *et al.*, 2016).

¹⁷As 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

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 to mitigate their activity levels (Corollary 3.5). When liability rules are altered, (perfect) markets will react. Under MRL rules, manufacturers are likely to implement technical and contractual solutions to transmit at least some of the cost of their residual liability downstream to operators. This will then incentivize operators toward 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' activity level. In a competitive market, we 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 optimal activity levels. As discussed in Corollary 3.3, this will in turn incentivize high-usage operators to purchase safer technology entailing lower activity-level operation costs.

The proposed MRL 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 manufacturers to operators under the MRL rule, no inverse mechanisms can be constructed as easily to transmit incentives upstream from operators to manufacturers.¹⁸ 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 preferable because the manufacturer can affect downstream incentives, whereas operators will have difficulty affecting upstream incentives.

We conclude with three remarks.¹⁹ The first concerns the victims' activity levels. If operators and victims are contractually related (e.g. the operator is using a self-driven car as a taxi), the liability cost would be included in the service cost, and also the activity level of the victim would be mitigated. However, in the absence of any market or contractual relationship, the potential victims may undertake excessive activity levels. Thus, the approach is second-best, and reflects the sensible argument that manufacturer safety – and, in turn, the operators' activity levels – are more important factors in preventing accidents, compared to the victims' activity levels. Generally, there is always one unregulated choice, and the optimal rule turns on the empirical question of which is the least important in determining risk.²⁰

The second remark concerns the interpretation of the operators' 'activity'. In standard products liability models (Hay and Spier, 2005; Polinsky, 1980), the users' 'activity' is generally interpreted in terms of *output*, i.e. as the number of units of the product that is purchased. Hence, manufacturers' and consumers' (victims') activity are coincident, and are determined by market equilibrium. This is the source of the so-called 'irrelevance result' in standard models, which maintains that the optimal

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.

¹⁸Market 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.

¹⁹We are indebted to an anonymous referee for insightful comments on the discussion that follows.

²⁰It can also be possible that manufacturers – faced with residual liability – 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 accidents, a different allocation of residual liability could be considered.

output will arise in such models regardless of the assignment of liability – provided that risk is correctly perceived by consumers. In our model, the users’ ‘activity’ is instead interpreted in terms of *usage*, i.e. as usage level per unit purchased. This definition follows the standard accident models (Landes and Posner, 1987; Shavell, 1980, 1987), and it is often related to consumers’ ‘care’. This different interpretation does not alter our conclusions. Indeed, under MRL rule, the price reflects the full cost per unit (including residual liability), and so operators would buy the efficient number, assuming that they correctly perceive the risk.

As third note, we shall point out that the hypothesized percolation effect of residual liability on activity levels (Corollary 3.5) does not undermine the manufacturer’s incentives to produce safer robots (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, 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 MRL rules, the victim’s incentives to prove the operator’s negligence may be reduced. 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. MRL rules 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. This would limit their liability exposure in case of accidents caused by the operator’s negligence. For example, evidence technology such as black boxes or dash-cams could be installed in self-driving cars; or, a decision log can be installed in a surgical robot to record the robot’s choices in 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.²¹ 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 (Guerra and Parisi, 2020). When the manufacturer’s residual liability operates under rules of contributory or comparative negligence (Figure 1), manufacturers would also have incentives to monitor victims’ care levels, since under these rules finding negligence by victims would equally shield manufacturers from liability.

²¹In robot torts, the proof of the operator’s negligence could be more difficult for victims 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.

5. Final remarks

This paper provides a novel economic analysis of robot-generated accidents. We propose an ‘MRL’ rule – 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. We show that this rule provides a second-best efficient set of incentives, accomplishing four objectives: incentivizing (1) efficient levels of human care by operators and victims, (2) efficient activity levels in the use of robots; (3) efficient R&D investments for the development of safer robots; and (4) adoption of safer robots in the marketplace. We further show the positive effects of our rule on the incentives to produce evidence for the accurate adjudication of robot torts.

At present, this rule cannot be observed in practice and we can only offer theoretical conjectures on its operation. The effectiveness of this rule *vis-à-vis* other liability regimes in creating optimal incentives may deserve some experimental and empirical investigation. Future scholarship should extend our basic model to consider operators’ risk aversion; imperfect markets; imperfect information on parties’ care levels; and the liability of the programmer as distinct from that of the manufacturer.²² Future scholarship might also consider how market relationships between operators and victims could align activity-level incentives also for victims; and how incentives of operators and manufacturers would change if robots were held responsible for accidents under a negligence standard, comparing their independently performed activities and safety level to those of a reasonable person or ‘reasonable robot’.

To conclude, it is worth remarking that any liability regime and legal reform have institutional dimensions and consequences. In our companion paper (Guerra *et al.*, 2021), we discuss the prospect offered by the conceptualization of robots as institutions by granting them electronic legal personhood. In this paper, we have offered a general framework for identifying the optimal liability regime for robots – general enough to be potentially implemented in various institutional contexts. To isolate the effects and derive results with predictive power, our analysis necessarily assumes away many institutional elements – whose investigation represents the natural extension of our research.

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**(1): 127–171.
- Ben-Shahar, O. (2016), ‘Should Carmakers be Liable When a Self-Driving Car Crashes’, *Forbes*, published online on Sep 22, 2016, available at <https://www.forbes.com/sites/omribenshahar/2016/09/22/should-carmakers-be-liable-when-a-self-driving-car-crashes/> (last accessed Nov 2021).
- 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. (2015), ‘Robotic Prostheses as Products Enhancing the Rights of People with Disabilities. Reconsidering the Structure of Liability Rules’, *International Review of Law, Computers & Technology*, **29**(2–3): 116–136.
- Bertolini, A. (2020), ‘Artificial Intelligence and Civil Liability’, *Bruxelles: European Parliament – Committee on Legal Affairs*, **608**: 1132.
- 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*, **51**(27): 243–284.
- Bertolini, A., P. Salvini, T. Pagliai, A. Morachioli, G. Acerbi, F. Cavallo, G. Turchetti and P. Dario (2016), ‘On Robots and Insurance’, *International Journal of Social Robotics*, **8**(3): 381–391.
- 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.
- Casey, B. (2019), ‘Robot Ipsa Loquitur’, *Georgetown Law Journal*, **108**(2): 225–286.
- Casey, B. and M. A. Lemley (2019), ‘You Might Be a Robot’, *Cornell Law Review*, **105**(2): 287–362.
- 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**(2): 191–320.

²²We are grateful to Eric Talley and two anonymous referees for these insightful suggestions.

- 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*, **170**(4): 571–598.
- De Chiara, A., I. Elizalde, E. Manna and A. Segura-Moreiras (2021), 'Car Accidents in the Age of Robots', *International Review of Law and Economics*, **68**(in progress): 106022.
- Epstein, R. A. (2021), 'Liability Rules for Autonomous Vehicles', *Journal of Entrepreneurship and Public Policy*, **10**(2): 218–234.
- European Commission (2020), 'Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics (COM/2020/64 final)', Available at <https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:52020DC0064> (accessed 07 October 2021).
- 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.
- Grady, M. (1987), 'Why are People Negligent Technology, Nondurable Precautions, and the Medical Malpractice Explosion', *Northwestern University Law Review*, **82**(2): 293–334.
- Grady, M. F. (2009), 'Unavoidable Accident', *Review of Law & Economics*, **5**(1): 177–231.
- Guerra, A. and F. Parisi (2020), 'Investing in Private Evidence', Available at SSRN https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3596029.
- Guerra, A., F. Parisi and D. Pi (2021), 'Liability for Robots I: Legal Challenges', *Journal of Institutional Economics*, published online. doi:10.1017/S1744137421000825.
- Hay, B. and K. E. Spier (2005), 'Manufacturer Liability for harms Caused by Consumers to Others', *American Economic Review*, **95**(5): 1700–1711.
- Kovač, M. (2020), *Judgement-Proof Robots and Artificial Intelligence: A Comparative Law and Economics Approach*, London, UK: Palgrave Macmillan Springer Nature.
- Landes, W. M. and R. A. Posner (1987), *The Economic Structure of Tort Law*, Cambridge, Massachusetts, US: Harvard University Press.
- Lemley, M. A. and B. Casey (2019), 'Remedies for Robots', *The University of Chicago Law Review*, **86**(5): 1311–1396.
- Luppi, B., F. Parisi and D. Pi (2016), 'Double-edged Torts', *International Review of Law and Economics*, **46**(June 2016): 43–48.
- Miceli, T. J. (1997), *Economics of the Law: Torts, Contracts, Property, Litigation*, Oxford: Oxford University Press.
- Mot, J. D. and B. Depoorter (2011), 'Technology and Torts: Memory Costs, Nondurable Precautions and Interference effects', *International Review of Law and Economics*, **31**(4): 284–290.
- Polinsky, A. M. (1980), 'Strict Liability vs. Negligence in a Market Setting', *The American Economic Review*, **70**(2): 363–367.
- Polinsky, A. M. and Y.-K. Che (1991), 'Decoupling Liability: Optimal Incentives for Care and Litigation', *RAND Journal of Economics*, **22**(4): 562–570.
- Roe, M. (2019), 'Who's Driving That Car: An Analysis of Regulatory and Potential Liability Frameworks for Driverless Cars', *BCL Rev*, **60**(1): 315–347.
- Shavell, S. (1980), 'Strict Liability versus Negligence', *The Journal of Legal Studies*, **9**(1): 1–25.
- Shavell, S. (1987), *Economic Analysis of Accident Law*, Cambridge, Massachusetts: Harvard University Press.
- Shavell, S. (2020), 'On the Redesign of Accident Liability for the World of Autonomous Vehicles', *The Journal of Legal Studies*, **49**(2): 243–285.
- Singh, R. (2006), 'On the Existence and Efficiency of Equilibria Under Liability Rules', *NBER Working Paper 12625*.
- Talley, E. (2019), 'Automotorts: How Should Accident Law Adapt to Autonomous Vehicles? Lessons from Law and Economics', Available at <https://www.hoover.org/sites/default/files/ip2-19002-paper.pdf> (last accessed 07 October 2021).
- Weston, S. B. (1963), 'Contributory Negligence in Product Liability', *Cleveland-Marshall Law Review*, **12**(3): 424–436.

Appendix A

Proof of Proposition 3.1

Proving this proposition requires us to show that (1) the operator 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 efficient care by both parties is a Nash equilibrium for any $\theta \in [0, 1]$. This a well-known result (e.g. Carbonara *et al.*, 2016), hence the proof is omitted for brevity.

Regarding point (2), in equilibrium, the manufacturer minimizes $r + (w^*z^*/r)p(x^{**}, y^{**})L$. The minimization problem yields $r^*:(wz/r^2)pL = 1$, which is equal to the socially optimal level as defined in equation (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. \square

Proof of Proposition 3.4

Proving this proposition requires us to show that the operator and the victim may have incentives to undertake excessive activity levels. In equilibrium ($x^* = x^{**}$ and $y^* = y^{**}$), the operator 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 operator 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, under rules of MRL, it follows that $z^* > z^{**}$ and $w^* > w^{**}$. This misalignment of activity-level incentives is necessary to incentivize optimal $r^* = r^{**}$. \square