The Korean Economic Review
Volume 40, Number 1, Winter 2024, 49-75.
DOI : 10.22841/kerdoi.2024.40.1.002

Law and Economics of Artificial Intelligence: Optimal Liability Rules for Accident Losses Caused by Fully Autonomous Vehicles*

Jeong-Yoo Kim**

We examine the optimal liability rule in accidents involving fully autonomous vehicles. In cases where enforcing due activity is not feasible, it is socially optimal to apply the strict liability rule to the human operator determining the activity level and to apply the negligence rule to the manufacturer and the victim who select care levels under contributory or comparative negligence in the unilateral activity case. Under the joint and several liability rule, both the manufacturer and the victim exercise due care, contingent on regulating the manufacturer's liability share sufficiently high, and the human operator assuming the remaining risks chooses the socially optimal activity level maximizing the social net benefit. Conversely, if due activity enforcement is possible, an alternative liability rule proves optimal. Under this rule, the human operator engages in efficient activity to comply with the activity standard, the manufacturer exercises efficient care to meet the care standard, and the victim assumes residual liability so as to be induced to take efficient care. Notably, this liability rule achieves the social optimum, even in bilateral activity cases where both the human operator and the victim engage in activity. Our results diverge from previous findings suggesting that achieving the social optimum involves using public sanctions, such as paying a fine to the state.

JEL Classification: K13
Keywords: Artificial Intelligence, Autonomous Vehicles, Due Activity, Joint and Several Liability, Liability Rule, Residual Liability

I. Introduction

The integration of artificial intelligence (AI) is reshaping traditional human

Received: Nov. 25, 2022. Revised: Jan. 30, 2023. Accepted: March 15, 2023.
* This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2022S1A5A2A0304932311).
** Professor, Department of Economics, Kyung Hee University, 1 Hoegidong, Dongdaenumku, Seoul 130-701, Korea, (Tel & Fax) +82-2-961-0986, Email: jyookim@khu.ac.kr

intelligence across various domains. Unlike their human counterparts, robots operate ceaselessly, devoid of the need for rest or vacations, and they are immune to the spread of infectious diseases. This undeniable shift positions robots as indispensable contributors to future production processes.

Among the sectors benefiting from AI technologies, autonomous driving stands out prominently. The desire for enhanced safety prompts drivers to seek innovative solutions. A notable advantage emerges when robots take the wheel, as they adhere precisely to programmed traffic rules, showcasing superior driving accuracy compared with humans. Additionally, their capacity to swiftly respond to unforeseen situations surpasses human capabilities. Most significantly, the liberation of humans from the task of driving allows for valuable opportunities to rest or attend to urgent business matters while AI technologies handle the driving responsibilities.

An essential concern arising in the era of autonomous driving pertains to determining liability for losses resulting from car accidents.[1] This issue is closely tied to the level of autonomy, as classified by the Society of Automotive Engineers (SAE) into six levels.[2] In the initial two levels (0–2), human drivers retain control over driving environments, with the system managing only specific driving functions. Level 3 introduces a shift, where the system takes over driving responsibilities, and the human driver intervenes solely in unexpected circumstances. At level 4, human intervention is eliminated entirely, as the system controls driving under all circumstances. Finally, at level 5, the system guides the autonomous vehicle seamlessly to its destination without any involvement from human drivers. Determining liability for accidents is intricately linked to these distinct levels of autonomy.

In the context of autonomous vehicles operating at levels 0–2, it is reasonable to attribute liability to human drivers for losses resulting from car accidents because they maintain control over the system.[3] This responsibility aligns with their role in overseeing the driving environment. Similarly, at level 3, where human drivers intervene in driving during unexpected circumstances, their liability may extend only to losses arising from accidents in such specific situations. However, as

[1] Interestingly, the question of whether different standards of care should apply to human drivers compared with AI drivers was already addressed by the court in the 1950s, a time when autonomous vehicles did not exist. In the case of Arnold v. Reuther, Mrs. Arnold brought a negligence claim against Mr. Reuter, whose car collided with her when she darted across the street. The court dismissed Mrs. Arnold's argument that Mr. Reuter had the last clear chance to avoid the accident, asserting that his efforts were sufficient to avoid liability given his human limitations. This implies that, had the vehicle been autonomous, it might have been held to a higher standard of care. For more details on the case, refer to Vladeck (2014).

[2] See SAE J3016 (2021).

[3] In case of system faults, human drivers may pursue contribution claims against the manufacturer or system maker.

autonomy reaches level 4, holding human operators liable for accident-related losses becomes challenging to justify.[4] Vehicles classified as level 4 or above are termed fully autonomous, whereas those at level 3 or below fall under the category of semi-autonomous vehicles.

Currently, in most countries, human drivers or operators bear liability for losses resulting from accidents. As we approach the era of level 4 autonomous vehicles, there is a pressing need for legislative amendments to align with this evolving technology. The Korean government has taken a proactive stance, explicitly stating in the Road Map 2.0 on Regulatory Innovation for Autonomous Vehicles, announced in January 2022, that manufacturers should be liable for losses stemming from accidents involving autonomous vehicles of levels 4–5.[5] However, our paper contends that assigning sole liability to the manufacturer for all losses due to accidents caused by autonomous vehicles at level 4 or above may be inefficient.

A critical aspect explored by scholars in law and economics is determining the socially optimal rule of tort liability. Two competing rules are considered: the strict liability rule and the negligence rule. Under the strict liability rule, an injurer is obligated to cover all accident losses they cause, irrespective of their level of care. By contrast, the negligence rule holds an injurer liable for accident losses only if they were negligent, i.e., if their care level was below what is considered due care.

The efficiency of both rules is widely acknowledged if courts have sufficient information to appropriately set the negligence standard, unless expected accident losses depend on factors such as the victim's care level or the injurer's activity level in addition to the injurer's care level. Shavell (1980) addresses both scenarios: one where the accident probability depends on the victim's care level and another where it depends on the injurer's activity level. His main conclusions are that (i) in the former case, the outcome will be socially optimal under some form of the negligence rule,[6] whereas it will not be under strict liability because the victim may neglect to take any care, and (ii) in the latter case, strict liability creates the efficient incentive to take care and activity, whereas the negligence rule induces too much activity because the injurer is not held liable for accident losses under the negligence rule simply because they engage in too much activity.

---

[4] In Forbes, Ben-Shahar (2016) made a similar conclusion that "There will be no drivers to blame, and the only remaining culprit would be the technology."

[5] Certain vehicle manufacturers, including Volvo and Mercedes, have declared their commitment to covering expenses for any injury or damage resulting from their fully autonomous cars.

[6] In broad terms, the three forms of the negligence rule are simple negligence, contributory negligence, and comparative negligence. Under simple negligence, the injurer is held liable if and only if they fail to take due care, regardless of the victim's care level. In contributory negligence, the injurer is liable if and only if they fail to take due care, and the victim takes due care, meaning the injurer is negligent, and the victim is not contributorily negligent. Comparatively, under comparative negligence, the injurer is fully liable if only they fail to take due care. However, they are partially liable in proportion to their relative negligence if both parties fail to take due care.

However, what constitutes the optimal liability rule remains relatively unclear when the accident probability depends on both the victim's care level and activity level as well as the injurer's care level and activity level. Note that neither the strict liability rule nor the negligence rule can achieve social optimum in this scenario. Under the strict liability rule, the injurer assumes residual liability, prompting them to adopt efficient care and activity levels. However, the other party (victim) tends to take too little care. Conversely, under the negligence rule, the victim bears residual liability, leading to the victim taking efficient care, but the injurer engaging in excessive activity. Various approaches exist to address this dilemma. Vickrey (1968) proposes congestion pricing to internalize the accident externalities of driving (activity), and Green (1978) suggests the double strict liability rule, holding both the injurer and the victim fully liable by having the injurer pay damages to the state. Additionally, Edlin (1998) proposes a per-mile insurance premium high enough for a driver to cover the expected value of injuries to both themselves and others.

The case of fully autonomous vehicles (level 4 or above) possesses a distinctive characteristic. The human operator cannot be negligent because they do not control the system while driving; it is the system itself that drives. However, the human operator does exert influence by selecting the destination, essentially determining the activity level. If the activity level also impacts the accident probability, autonomous car accidents involve multiple tortfeasors—the system and the human operator.

In this paper, we examine accidents involving level 4 autonomous vehicles, asserting that in the unilateral activity case where the accident probability is independent of the victim's activity level, if direct regulation of the activity level is not feasible by setting a due activity level, it is socially optimal to apply the negligence rule to the manufacturer and the victim who select care levels.[7] Simultaneously, the strict liability rule should be applied to the human operator responsible for determining the activity level. Under the joint and several liability rule, the manufacturer adheres to due care if their liability share is set high, and the human operator, assuming all remaining risks, selects the socially optimal activity level maximizing the net social benefit. However, if the strict liability rule is applied to the human operator, the victim may lack incentive to take efficient care, as injurers bear all responsibility for accident losses regardless of the victim's care level. To counter this adverse incentive, a contributory negligence or comparative negligence rule is necessary, where injurers are liable only for the damage amount discounted in proportion to the victim's negligence. Alternatively, if the manufacturer bears primary liability under the negligence standard, and the human operator bears residual liability, instead of applying joint and several liability, the manufacturer takes due care, the human operator engages in efficient activity to

[7] The care level of the manufacturer can be construed as the precision or accuracy of the system.

internalize all accident externalities, provided the victim exercises due care to avoid liability. This rule, termed a human operator residual liability (HRL) rule, can achieve social optimum without any conditions.[8] Unfortunately, the HRL rule cannot induce social optimum in the bilateral activity case where the accident probability depends on both the victim's and the injurer's activity levels.

Some may argue that monitoring a human operator's activity level is technically feasible in the context of fully autonomous vehicles. Therefore, the imposition of due activity on human operators is conceivable. However, enforcing this in reality poses challenges, as legal authorities cannot ascertain each individual's private information regarding their valuation from the activity, which is crucial for determining due activity. If, however, due activity can be effectively enforced, the opposite liability rule becomes optimal. Under this rule, the human operator engages in efficient activity to adhere to the activity standard, and the manufacturer takes efficient care to meet the care standard. In essence, broadly defined negligence rules are applied to the joint injurers, and the victim assumes residual liability, compelling them to adopt efficient care. This rule is termed a victim residual liability (VRL) rule. This liability rule can attain social optimum, even in the bilateral activity case where both the human operator and the victim engage in activity.

Recently, a growing body of literature has emerged, focusing on the optimal liability rule for autonomous vehicles. Vladeck (2014) employs a traditional principal-agent analysis to advocate for the apportionment of liability among human drivers, manufacturers, and consumers.[9] Abraham and Rabin (2019) put forward the concept of manufacturer enterprise responsibility (MER), which entails auto manufacturers being responsible for all injuries stemming from the operation of fully autonomous vehicles. Logue (2019) extends a similar proposal to semi-autonomous vehicles, contributing to this discourse.

In a more formal analysis, Talley (2019) employs the standard bilateral precaution framework to scrutinize interactions between algorithmic and human decision-makers. He explores the efficacy of various liability structures in distributing risks during a transition period, where both human and autonomous

[8] Under the joint and several liability rule, if the manufacturer is not negligent, the human operator assumes residual liability. However, if the manufacturer is found negligent, both the manufacturer and the human operator become jointly liable. Therefore, the characterization of the human operator as a true residual liability bearer may be nuanced in this context.

[9] Vladeck (2014) puts forth the argument, using the Toyota case of "sudden acceleration," that consumers may encounter challenges in seeking compensation for auto accident losses from manufacturers under current product liability laws, even if the vehicle involved is fully autonomous. This difficulty arises from the challenge of proving product defects. In the Toyota case, experts struggled to identify a specific design or manufacturing defect causing the uncontrolled acceleration. Similarly, the case of the Tesla autopilot crash in 2016 faced a similar situation, with the investigation concluding that despite the crash, there were no defects warranting a recall.

vehicles may coexist. His findings indicate that certain negligence-based rules can achieve efficient outcomes. However, Talley's analysis does not delve into the impact of the liability rule on the activity level of the human driver or the manufacturer's incentive to invest in the safety of autonomous vehicles. Guerra et al. (2021) propose the manufacturer residual liability (MRL) rule, distinct from MER, where the manufacturer is liable for the entire accident losses of autonomous vehicles. Under MRL, human operators and victims bear losses due to negligence, and manufacturers are residually liable for non-negligent accidents. Applying the negligence rule to operators and victims ensures compliance with negligence standards, and the manufacturer, being residually liable, makes optimal investments similar to under strict liability. Although this liability rule makes sense in a world of level 3 or below, in a world of level 4 or 5, human operators choose activity levels which are not generally optimal under this rule, as Guerra et al. (2021) note by writing "allocating residual liability on manufacturers may lead to excessive activity levels for both operators and victims."[10] Shavell (2020) proposes a novel form of strict liability wherein damages would be paid to the state instead of the victim, aligning with the insight of Green (1976). De Chiara et al. (2021) consider a perfectly competitive market for traditional vehicles and a monopoly market for autonomous vehicles.[11] Despite addressing a similar issue, their model is more comprehensive, encompassing pricing decisions, investment decisions of the monopolistic autonomous vehicle manufacturer, and potential decisions of human drivers to purchase fully-autonomous vehicles. However, they remain silent on the liability share between joint injurers (the human driver and the autonomous vehicle manufacturer), a key feature in our model.

In Sections 2 and 3, we consider the case of unilateral accident and the case of bilateral accident respectively. In Section 4, we discuss the possibility of applying due activity. We briefly discuss possible extensions in Section 5. We make concluding remarks in Section 6.

[10] To address the incentive for excessive activity among human operators, the authors propose a price mechanism where manufacturers charge human operators a fee.

[11] Several papers discuss liability rules in a market setting. Calabresi (1961) is acknowledged as the first to suggest that placing the full costs of accidents on the firm's side in a monopolized industry may not be socially desirable. Shavell (1980) and Polinsky (1980) demonstrate, in a perfectly competitive environment, that the strict liability rule is efficient, whereas the negligence rule is not efficient in the case of unilateral accidents. This paper, however, does not consider a market setting, aiming to isolate the care issue from the pricing issue by implicitly assuming that the price of autonomous vehicles is regulated.

II. Unilateral Accident

A potential defendant operates a level 4 or above autonomous vehicle, posing the risk of a car accident to a potential plaintiff who is the victim.[12] Here, we consider that the human operator's care level in driving the car is irrelevant to the accident risk due to the vehicle's advanced autonomy. This section concentrates on scenarios where the risk of an accident is solely influenced by the quality of the vehicle's system software (the care level of the car manufacturer)[13] and the extent of the human operator's activity (their choice of activity). Thus, our focus is exclusively on unilateral accidents where the victim's care level does not impact the accident probability.[14]

The following notations are used throughout the paper:

$x =$ the care level of the manufacturer $(x \geq 0)$
$a =$ the activity level of the human operator $(a \geq 0)$
$\overline{x} =$ the level of due care
$p(a,x) =$ the probability of an accident $(p_a(a,x) > 0, p_{aa}(a,x) > 0, p_x(a,x) > 0, p_{xx}(a,x) > 0, p_{ax}(a,x) = 0$ for all $x, a \geq 0)$
$c_M(x) =$ the manufacturer's cost of preventing an accident $(c_M'(x) < 0, c_M''(x) > 0)$
$L =$ the losses caused by a car accident $(L > 0)$
$B_H(a) =$ the private benefit to a human operator engaging in their activity at level a $(B_H'(a) > 0, B_H''(a) < 0)$

For convenience, we employ the convention that a high x represents less care (or more carelessness).[15] Thus, x can be interpreted as a level of carelessness. Note also that $p(a,x)$ and L are both defined for one accident, implicitly assuming that an accident occurs at most once.[16]

Given the multiple defendants, the joint and several liability rule can be applied.

---

[12] Potentially numerous victims are heterogeneous in terms of their cost of taking care, their activity type, and their mobility. For example, victims can be drivers of non-autonomous vehicles, passengers of fully autonomous vehicles, or pedestrians. Given the complexity of considering multiple potential victims of various types, this analysis simplifies the scenario by assuming the presence of a representative single victim.

[13] In this context, by a manufacturer, we refer to a firm responsible for producing both the hardware and system software of autonomous vehicles. Although, in reality, a distinction between a system software supplier and an automobile manufacturer exists, for the sake of simplicity, we assume that the manufacturer provides both hardware and software.

[14] We do not consider the purchasing decision of a human injurer, implicitly assuming that autonomous vehicles are necessary commodities.

[15] This convention is to treat x and a symmetrically, following examples such as Kornhauser and Revesz (1989).

[16] If $p(a,x) = ap(x)$, we may interpret $p(x)$ as an accident probability per activity and L as the accident loss per occurrence. Our interpretation includes this case.

Under the joint and several liability, if an accident occurs, the victim may proceed jointly against both defendants or may choose to recover all damages from only one of them.

We assume that the manufacturer chooses the care level and the human operator chooses the activity without observing the other's choice. Henceforth, we denote the human operator by D_H, the manufacturer of the vehicle by D_M, and the victim by P.

The accident probability $p(a,x)$ can be interpreted as a production function using a and x as inputs. Subsequently, p_a and p_x have the interpretation of marginal productivity of an activity level and a care level. We assume that the two inputs are independent, i.e., $p_{ax} = 0$, implying that the marginal effect of a care level in reducing the accident probability does not depend on the activity level. This also means that the accident probability function $p(a,x)$ is additively separable. The signs of p_{aa} and p_{xx} reflect the increasing marginal productivity of the activity and the care. The first effect can also be called the depreciation effect. As a increases, the vehicle depreciates, leading to a larger marginal productivity, i.e., an increased marginal impact on the accident probability.

2.1. Social Optimum

Calabresi (1970) proposes the widely accepted efficiency goal of tort law, which is the minimization of the sum of precaution and accident costs, i.e.,

$$C(x) = p(x)L + c_M(x). \tag{1}$$

However, considering the human operator's activity that benefits them, the efficiency goal should be modified to incorporate this benefit.

The general social welfare is defined by the social benefits minus social costs, where the latter is the sum of accident costs and accident-preventing costs. Let W be social welfare. We then have

$$W(a,x) = B_H(a) - C(a,x) = B_H(a) - (p(a,x)L + c_M(x)). \tag{2}$$

Social optimum is attained by maximizing W. Let (a^*, x^*) be the socially optimal activity level and care level. It is then socially optimal for D_H and D_M to choose a^* and x^* respectively to maximize social welfare, i.e.,

$$\max_{a,x} W(a,x) = B_H(a) - (p(a,x)L + c_M(x)). \tag{3}$$

The first-order conditions can be written as

$$W_1 = B'_H(a^*) - p_a(a^*)L = 0\ , \tag{4}$$
$$W_2 = -p_x(x^*)L - c'_M(x^*) = 0\ . \tag{5}$$

Equation (4) is conventionally interpreted as follows: The socially optimal activity level should strike a balance between the social marginal benefits arising from an escalation in the driving level and the marginal cost associated with increased driving due to a rise in accident probability. Likewise, Equation (5) bears a comparable interpretation: The socially optimal care level for the manufacturer should find balance between the social marginal benefits derived from heightened care, resulting in a reduction in accident probability, and the marginal cost incurred by increased diligence.

The second-order condition of this optimization is satisfied because $|H_W| \equiv -(B''_H - p_{aa}L)(p_{aa}L + c''_M) > 0$, where H_W is the Hessian matrix of W.

2.2. Strict Liability

Considering the strict liability rule, once an accident occurs, D_H and D_M are jointly and severally liable for the damage L, for any x and a.

Let $\theta_i \in [0,1]$ be the liability share of $D_i, i = H, M$ when an accident occurs. The joint and several liability rule requires that $\theta_H + \theta_M = 1$ because the sum of the damages each defendant pays to the victim cannot exceed the entire amount awarded under the rule.

Let a^S and x^S be the equilibrium activity level of D_H and the equilibrium care level of D_M respectively under the strict liability rule. In addition, let $C_i^S(a,x)$ be the expected cost that D_i bears under the strict liability rule. If $\theta_H, \theta_M \in (0,1), D_H$ chooses *a* to maximize

$$\max_a \phi(a, x^S) \equiv B_H(a) - C_H^S(a, x^S) = B_H(a) - p(a, x^S)\theta_H L\ , \tag{6}$$

and D_M chooses x to minimize

$$\min_x C_M^S(a^S, x) = c_M(x) + p(x, a^S)\theta_M L \equiv \psi(a^S, x)\ . \tag{7}$$

The first-order conditions require

$$\phi_a \equiv B'_H(a^S) - p_a(a^S)\theta_H L = 0\ , \tag{8}$$
$$\psi_x \equiv c'_M(x^S) + p_x(x^S)\theta_M L = 0\ . \tag{9}$$

Let us compare private optimum (a^S, x^S) given by (8) and (9) with social optimum (a^*, x^*) given by (4) and (5). First, a^* and x^*, which are the solution for (4) and (5), do not satisfy (8) and (9), implying that the strict liability rule is inefficient. Second, $a^S > a^*$ or $x^S > x^*$, i.e., D_H takes too much activity or D_M takes too little care.

Proposition 1. *Under the strict liability rule, (i)* (a^*, x^*) *cannot be an equilibrium, and (ii)* D_H *chooses an excessive activity level* $(a^S > a^*)$ *or* D_M *takes too little care* $(x^S > x^*)$.

It depends on the sizes of θ_H and θ_M whether $x^S > x^*$ or $a^S > a^*$. If $\theta_M = 1$, D_M takes efficient care $x^S = x^*$. However, if D_M chooses x^* and takes full liability, D_H takes no liability at all under the joint and several liability. Therefore, they take excessive activity level, i.e., use their vehicle for moving too much. If $\theta_H = 1$, then $\theta_M = 0$, implying that D_M has no incentive to take care. Thus, the strict liability rule cannot achieve social optimum in any case.

2.3. Negligence

A human operator cannot have the legal standard of due care because they only choose the activity level for a fully autonomous vehicle. Hence, we consider the case that the negligence rule is applied only to the manufacturer, whereas the strict liability rule is applied to the human operator.[17] That is, if the car manufacturer is negligent, the human operator and the manufacturer share liability with their respective liability share θ_H and θ_M. However, if the manufacturer is proven to comply with the due care, the accident loss has to be fully covered by the human operator to which the strict liability rule is applied. In this case, each liability share becomes $\theta_H = 1$ and $\theta_M = 0$.

Let a^N and x^N be the equilibrium activity level of D_H and the equilibrium care level of D_M respectively under this rule. Let $\overline{x} = x^*$. Given that D_M takes the due care $x^N = \overline{x} = x^*$, the rule implies that $\theta_H = 1$ and $\theta_M = 0$ because D_M is not negligent. Therefore, D_H chooses the activity level to maximize

$$\max_a \phi^N(a, x^N) = B_H(a) - C_H(a, x^N) = B_H(a) - p(x^N, a)L \,, \tag{10}$$

[17] Some contend that with fully autonomous vehicles capable of monitoring mileage, the activity level of the human operator becomes verifiable, suggesting its potential inclusion in a due care standard. However, we do not subscribe to the notion that verifiable activity levels should necessarily be incorporated into the standard of due care. In our view, even if the activity level is verifiable, it would be impractical and scarcely acceptable to impose liability on a human operator solely for choosing an activity level that exceeds a legally allowed threshold, i.e., driving excessively. Further discussion on this matter is available in Section 4.1.

and the first-order condition requires

$$\phi_a^N(a,x^N) = B_H'(a) - p_a(a)L = 0 \ . \tag{11}$$

Equation (11) is identical to Equation (4), implying that $a^N = a^*$. On the contrary, given that $a^N = a^*$, $x^N = x^*$, if θ_M is large enough.

Proposition 2. *Under the negligence rule, there exists* $\underline{\theta}_M \in (0,1)$ *such that for any* $\theta_M \geq \underline{\theta}_M$, D_H *chooses the efficient activity level and* D_M *chooses the efficient care level in equilibrium.*

This proposition implies that efficiency is not guaranteed under the negligence rule unless θ_M is very large. If θ_M is very small, that $x^N > x^*$ is possible, i.e., the manufacturer takes too little care because they do not expect to be liable very much. If θ_M is very large, the negligent manufacturer takes most of the liability. Accordingly, the manufacturer adopts the efficient level of care, prompting the human operator to engage in the efficient level of activity. This implies that efficiency necessitates the regulation of liability shares, with a greater share of liability placed on the manufacturer, under the joint and several liability with the negligence standard. This outcome appears to diverge from the findings of Landes and Posner (1980) and Kornhauser and Revesz (1989), who posit that under the joint and several liability rule, a negligence rule induces all potential tortfeasors to exercise efficient care regardless of liability shares. Note that our result is derived when the negligence rule is exclusively applied to the automobile manufacturer, whereas the strict liability rule is applied to the human operator.

2.4. Negligence with Human Operator's Residual Liability

To allocate liability, we can use the negligence rule coupled with various residual liability rules. Consider a liability rule that applies a negligence standard primarily to D_M and shifts residual liability to D_H. As a negligence rule is applied to D_M, the manufacturer takes due care $\overline{x}$, and because D_H bears residual liability, they internalize all the externality and thus choose the efficient activity level to maximize the social welfare. We refer to this rule as the HRL rule to contrast with the MRL rule by Guerra et al. (2021) established for accidents by autonomous vehicles of level 3 or below.

Proposition 3. *Under the negligence rule with the residual liability on* D_H, D_H *chooses the efficient activity level and* D_M *chooses the efficient care level in equilibrium.*

This proposition implies that the HRL rule is optimal in a world of fully autonomous vehicles of level 4 or above, whereas the MRL rule may be optimal in a world of semi-autonomous vehicles of level 3 or below.[18] This is mainly because the HRL rule can control the human activity level that would be excessive otherwise. The problem of excessive human activity can also occur for autonomous vehicles of level 3 or below. Guerra et al. (2021) discuss how to control the human activity level. This issue is revisited in Section 4.

Under this rule, social optimum is attained without any condition, whereas the joint and several liability attains social optimum with a specific condition on θ_M. The purpose of this rule is not to treat two defendants symmetrically. By applying the negligence rule first to the manufacturer, we have only two cases of liability shares. If the manufacturer chooses $x > \overline{x}$, the resulting liability shares are $(\theta_M, \theta_H) = (1,0)$. If the manufacturer chooses $x \le \overline{x}$, the liability shares are $(\theta_M, \theta_H) = (0,1)$. In other words, this rule allows only two pairs of liability shares among a continuum set of liability shares $(\theta_M, \theta_H) = (\theta, 1-\theta)$ for $\theta \in [0,1]$. As its core objective is to grant the manufacturer priority in choosing between these two pairs, in equilibrium, the manufacturer invariably selects $(\theta_M, \theta_H) = (1,0)$ by choosing $x = \overline{x}$ in equilibrium, leading the human operator to bear all the remaining liability. This outcome precisely aligns with the HRL rule.

III. Bilateral Accident

In this section, we consider a more comprehensive model of a bilateral accident where the accident probability hinges on both the care exercised by the victim and that on the defendant's side. For the case of bilateral accidents, we use the following additional notations:

$y =$ the care level of a victim $(y \ge 0)$
$\overline{y} =$ the due care level of the victim
$p(x, y, a) =$ the probability of an accident $(p_i(x, y, a) > 0$ and $p_{ii}(x, y, a) > 0$ for $i = a, x, y$ and for all $x, y, a \ge 0)$
$c_P(y) =$ the victim's cost of preventing an accident $(c_P'(y) < 0, c_P''(y) > 0)$

As a high value of x represents less care of a manufacturer, a high value of y

[18] Guerra et al. (2021) argue that the MRL rule is preferable to the HRL rule. This preference arises from the MRL rule's ability to convey residual liability incentives downstream from manufacturers to human operators through the price system, leveraging usage records. In contrast, the HRL rule lacks the capacity to transmit liability upstream. This argument holds merit when the liability rule is suboptimal, allowing the price mechanism to complement its incompleteness. However, in our model, the HRL rule stands as optimal, rendering a supplementary price mechanism unnecessary.

also means less care of a victim. The two care levels, x and y, can be substitutes $(p_{xy} < 0)$ or complements $(p_{xy} > 0)$, but a and y are independent $(p_{ay} = 0)$, just as a and x are independent. We assume a bilateral care but a unilateral activity in the sense that the victim is passive in activity.[19]

3.1. Social Optimum

The social optimum, (a^*, x^*, y^*), can be similarly defined as in the previous section. It is socially optimal for D_H, D_M, and P to choose a, x, and y, respectively, to maximize the social welfare defined by

$$W(a,x,y) = B_H(a) - C(a,x,y) = B_H(a) - (c_M(x) + c_P(y) - p(a,x,y)L). \tag{12}$$

The first-order conditions imply

$$W_1 = B_H'(a^*) - p_a(a^*)L = 0, \tag{13}$$
$$W_2 = -c_M'(x^*) - p_x(x^*, y^*)L = 0, \tag{14}$$
$$W_3 = -c_P'(y^*) - p_y(x^*, y^*)L = 0. \tag{15}$$

The equations carry conventional interpretations. Equation (13) suggests that the socially optimal activity level necessitates a balance between the social marginal benefit and the social marginal cost associated with elevating the activity level, achieved by increasing the accident probability. Equations (14) and (15) assert that D_M and P should select their care levels to strike a balance between the social marginal benefit, achieved by reducing the accident probability, and the corresponding social marginal cost.

The second-order condition is satisfied if the second cross derivative, $|p_{xy}|$, is not very large.

3.2. Contributory or Comparative Negligence

In cases of bilateral accidents, it is widely recognized that neither simple strict liability nor simple negligence leads to the first-best outcome when the accident

---

[19] In reality, potential victims engage in various levels of activity. However, for analytical simplification, we assume a single representative victim who exercises caution. For instance, consider a scenario where an autonomous vehicle collides with a pedestrian or crashes into a store due to malfunction. It is noteworthy that a human injurer's activity is crucial to the occurrence of an accident, whereas a victim's activity is not. Guerra et al. (2021) echo a similar sentiment, stating “the operator’s activity levels are more important factors in preventing accidents, compared to the victim’s activity levels.”

probability depends on both the care level and the activity level of the injurer. The rationale behind this observation can be explained as follows. First, the victim lacks any incentive to exercise care under the strict liability rule. Additionally, the negligence rule prompts the potential injurer to undertake excessive activity because it can only induce efficient care levels for both the injurer and the victim. When the injurer's activity influences the risk of accidents causing negative externalities, an injurer indifferent to the adverse effects of their activity on others engages in excessive activity.

However, if different individuals assume the roles of choosing care and activity levels, distinct liability rules can be applied to each, as discussed in Section 2. We consider a hybrid rule that applies the strict liability rule to the human operator who chooses the activity level and applies the negligence rule to all other legal parties under the joint and several liability rule.[20] In this scenario, the manufacturer and the human operator share liability, but if the manufacturer and the victim adhere to their respective negligence standards, the human operator bears sole responsibility for the entire loss.

Under this hybrid rule, the victim may not take enough care because the injurers are fully (jointly) liable as long as the strict liability rule is applied to D_H. To prevent the incentive of P to neglect care, we can incorporate contributory negligence or comparative negligence between injurers and the victim. If the victim is also negligent, the injurers may be jointly liable only for the damage amount discounted proportionally to the victim's negligence, whereas the victim is liable for the remaining damage amount. Such a modified comparative negligence rule serves to incentivize the victim to exercise care.

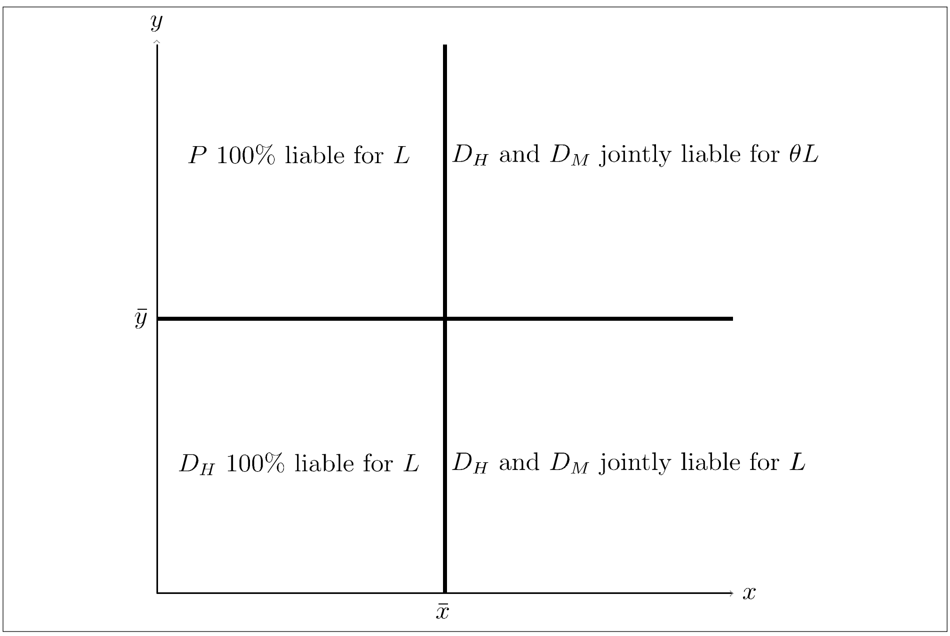
To elaborate, let $x-\overline{x}$ and $y-\overline{y}$ be the negligence of D_M and P respectively. We consider several cases. If $x-\overline{x}$, $y-\overline{y}>0$, i.e., both the manufacturer and the victim are negligent, they are liable in proportion to their negligence, i.e., the injurers are jointly liable for θL and the victim is liable for $(1-\theta)L$ where $\theta=\frac{x-\overline{x}}{x-\overline{x}+y-\overline{y}}$. Here, the denominator is the total negligence of the manufacturer and the victim, whereas the numerator of θ is the negligence of the manufacturer. If $y-\overline{y}\le 0$ and $x-\overline{x}>0$, i.e., only the manufacturer is negligent, injurers are jointly liable for the whole damage amount L by treating $y-\overline{y}$ as if $y-\overline{y}=0$. On the contrary, if $x-\overline{x}\le 0$ and $y-\overline{y}>0$, i.e., only the victim is negligent, we treat $x-\overline{x}$ simply as if $x-\overline{x}=0$. In this case, the victim is liable for the whole damage amount. Accordingly, D_H has no liability even under the strict liability rule. The absence of negligence on the part of the other injurer (manufacturer) creates a positive externality for the human operator. Finally, if

[20] This variant of the hybrid rule should be discerned from the hybrid (joint and several liability) rule, which differentiates between joint tortfeasors by establishing percentage thresholds of fault for defendants, as exemplified in California.

neither the manufacturer nor the victim is negligent (i.e., $x-\overline{x}, y-\overline{y} \leq 0$), the manufacturer and the victim are not liable, so D_H alone is fully liable under the joint and several liability rule. The taxonomy of all possible cases is described in Figure 1. A similar argument holds even when a contributory negligence rule, which is the limiting case of comparative negligence, is applied to D_M and P .

Proposition 4. *In a bilateral accident, under the joint and several liability with a hybrid rule whereby a strict liability rule is applied to* D_H *and a contributory or comparative negligence rule is applied to* D_M *and* P *with the due care levels* $\overline{x} = x^*$ *and* $\overline{y} = y^*$, *there exists* $\underline{\theta}_M \in (0,1)$ *such that for any* $\theta_M \geq \underline{\theta}_M$, *the activity and care levels of the injurers and victim are socially efficient in equilibrium.*

[Figure 1] Taxonomy: Joint and Several Liability with Comparative Negligence ($\theta = \frac{x-\overline{x}}{x-\overline{x}+y-\overline{y}}$)



An important difference from Proposition 2, its counterpart in the unilateral accident case, lies in the application of contributory or comparative negligence between the manufacturer and the victim.

The intuition behind this proposition is as follows: If the legal authority desires both the manufacturer and the victim to exercise due care, it should enforce the strict liability rule on the human operator to incentivize an efficient activity level. Under this strict liability rule, where the human operator bears 100% liability for losses resulting from the accident unless the victim is negligent, the human operator internalizes the negative externality and adopts a socially efficient activity level. In this context, the victim can opt for due care under contributory negligence or comparative negligence, as failing to do so would render them liable for the entire damage (in the case of contributory negligence) or a share of the damage proportional to their negligence (in the case of comparative negligence). With the assurance that the victim exercises due care, the manufacturer also does so if their liability share relative to the human operator's share is sufficiently substantial. Otherwise, the manufacturer would bear a significant portion of the liability. Consequently, in equilibrium, the human operator becomes solely liable for the entire damages, as both the manufacturer and the victim are exempted from any liability by adhering to due care.

3.3. Negligence with Human Operator's Residual Liability

If D_M and P bear primary liability under either comparative or contrib-utory negligence and D_H bears residual liability, D_M and P both take due care to avoid primary liability. D_H, who is the residual bearer, then takes the efficient activity level by internalizing all risks in equilibrium.

Proposition 5. *In a bilateral accident, if the comparative or contributory negligence rule is primarily applied to the manufacturer and the victim with the due care levels* $\overline{x} = x^*$ *and* $\overline{y} = y^*$*, respectively, and the human operator bears residual liability, the activity level of the human operator and the care levels of the manufacturer and victim are all socially efficient in equilibrium.*

Similar to the case of a unilateral accident, this HRL rule, which assigns residual liability to the human operator, proves superior to the joint and several liability rule by achieving the social optimum without any conditions. This scenario is essentially a special case of Proposition 4 in the sense that allocating primary liability on the manufacturer means $\theta_M = 1$ for his negligence.

Guerra et al. (2021) propose a liability rule with primary and residual liability, akin to ours, rather than a joint and several liability rule. However, their main finding contrasts sharply with ours, advocating for an MRL rule, whereas we support a HRL rule. This discrepancy arises from their primary focus on autonomous vehicles of level 3 or below. They also explore contributory and comparative negligence rules, aligning with our model.

De Chiara et al. (2021) also consider a hybrid rule similar to ours, but their liability rule differs in that it mandates the human defendant to pay a fine to the state when an accident occurs. In our model, where both the human operator and the manufacturer share liability for damages, efficiency can be achieved without resorting to public punishment for the human operator. The mechanisms employed by the two hybrid rules to incentivize the victim to exercise care differ slightly. In their model, the victim adopts efficient care because the human operator, subject to strict liability, pays damages (or a fine) to the state rather than the victim. Therefore, compensation for losses is contingent on the victim exercising due care. In our model, the contributory or comparative negligence rule compels the victim to adopt efficient care, as negligence would diminish or eliminate the court award.

IV. Discussions on Due Activity

The accident rate hinges on several factors, encompassing the care and activities of the legal parties involved. Nevertheless, our previous assumption posits the victim as a passive participant devoid of any activity level. That is, we assume that solely the human operator, guiding autonomous vehicles, determines the destinations, characterizing it as a unilateral activity case. Conversely, in a bilateral activity case, the probability of accidents is influenced not only by the activity level of the human operator but also by the activity level of the victims.

4.1. Unilateral Verifiable Activity

In the case of autonomous vehicles of level 4 or above, gauging the activity level of each human operator should not pose significant challenges. However, it does not automatically imply a straightforward application of due activity to human operators responsible for accidents caused by fully autonomous vehicles, for several reasons.

As we posit in Footnote 6, deeming a human operator liable for an accident solely because they chose an activity level surpassing the due activity level is scarcely acceptable. Although excessive carelessness may legitimately warrant liability, an abundance of driving does not seem to be a justifiable reason.

Moreover, the enforceability of due activity presupposes the ability of the legal authority to accurately compute the socially optimal level of activity, which depends on the human operator's benefit from the activity, $B_H(a)$. However, this benefit is subject to each individual's subjective valuation. Given the significant variation in private benefits across individuals, imposing a uniform due activity on all human operators, contrary to due care, is inappropriate. The key distinction between due care and due activity lies in the computability; due care can be derived from objectively measurable values, such as damage amount and accident probability, whereas due activity relies on knowing the private benefit, posing an information problem that hinders practical enforcement of a uniform due activity.

Nonetheless, if we assume that due activity is enforceable despite aforementioned difficulties in reality, i.e., the government can correctly compute a^* and set due activity $\overline{a} = a^*$, it can apply the negligence rule both to the human operator and the manufacturer by interpreting a as extended due care. Under joint and several liability, both the human operator and the manufacturer would adhere to due activity and due care, respectively. This gives rise to the VRL rule, where the residual liability borne by the victim compels efficient care, as they bear all associated risks. This rule obviates the need for the victim's contributory negligence to ensure the adoption of efficient care. Consequently, under this rule, the social optimum can be achieved regardless of whether contributory or comparative negligence is applied to the victim, as they are incentivized to take efficient care in any case.

4.2. Bilateral Activity

We now consider the bilateral activity case, where both the human operator and the victim engage in activity. Extending the logic from the unilateral activity case, let b be the activity level of a victim and $B_P(b)$ be the benefit of the victim from the activity level b. We can assume that b is unverifiable, whereas a is verifiable. The justification for this assumption is that a is the activity level of an autonomous vehicle that can keep track of its own activity level and b is the activity level of the victims including pedestrians or human-driven vehicles that generally cannot keep track of their activity levels. We also assume that the accident probability is $p(x,y,a,b)$ where $p_x, p_y, p_a, p_b > 0$ and $p_{xx}, p_{yy}, p_{aa}, p_{bb} > 0$. By abusing notation, let x^*, y^*, a^*, and b^* be the social optimum that maximizes the social welfare redefined by

$$W(a,b,x,y) = B(a) - C(a,b,x,y) = B_H(a) + B_P(b) \\ -[c_M(x) + c_P(y) + p(a,b,x,y)L]. \tag{16}$$

The first-order conditions imply

$$W_1 = B_H'(a^*) - p_a(a^*,b^*,x^*,y^*) = 0, \tag{17}$$

$$W_2 = B_P'(b^*) - p_b(a^*,b^*,x^*,y^*) = 0, \tag{18}$$

$$W_3 = -c_M'(x^*) - p_x(a^*,b^*,x^*,y^*)L = 0, \tag{19}$$

$$W_4 = -c_P'(x^*) - p_y(a^*,b^*,x^*,y^*)L = 0. \tag{20}$$

Let us consider private optimum under various liability rules. As we argue in the Introduction, the party bearing residual liability internalizes accident losses, leading to the adoption of efficient activity and care levels. This implies that if D_H is the residual bearer, then they choose the optimal activity level a^*; if P is the residual bearer, the victim takes the optimal activity b^*. However, both of the activity levels cannot be optimal.

However, if autonomous vehicles effectively track activities of human operators and the government possesses sufficient information about each individual's benefit function $B(a)$, this extra information can be leveraged to regulate a human's activity level. In this scenario, the legal authority may set the activity standard a and treat it as a form of due care. Thus, we can consider a liability rule where the human operator and the manufacturer primarily bear liability with due activity level and due care level, respectively, and the victim bears residual liability. This is the VRL rule, aiming to allocate residual liability to the party whose activity is unverifiable.

Under this rule, D_H chooses the activity level to maximize

$$\max_a \phi^{NR}(a,b,x,y) = \begin{cases} B_H(a), & \text{if } a \le \overline{a} \\ B_H(a) - p(x,y,a,b)L & \text{if } a > \overline{a}. \end{cases} \tag{21}$$

Setting $\overline{a} = a^*$ and $\overline{x} = x^*$ and assuming $\overline{y} = y^*$ and $b = b^*$, choosing the due activity $\overline{a}$ is optimal for D_H. As the victim bears residual liability, they internalize all the risk, adopting optimal care and activity levels. Therefore, the VRL rule can achieve social optimum in bilateral activity accidents if the due activity of the human operator is enforceable.

V. Discussions on Possible Extensions

Our model offers avenues for extension in various directions. In this section, we briefly explore some possibilities in connection with the existing literature.

5.1. Investment Decision

We may evaluate the manufacturer's investment decision in the context of De Chiara et al. (2021). Two approaches can be employed to integrate the investment into our model. One involves investing to decrease the manufacturer's cost of care, and the other entails investing to reduce the accident probability x, i.e., to increase the marginal productivity of x. Although De Chiara et al. (2021) consider on the former investment, our discussion centers on the latter.

Let I be the investment level of the manufacturer and $p(a,x,y,I)$ be a new accident probability function that incorporates this investment. Our hybrid liability rule cannot induce a social optimum because the manufacturer primarily prioritizes due care under the negligence rule, with limited concern for the investment level. Applying a strict liability rule to both the manufacturer and the human operator does not lead to a social optimum under joint and several liability, similar to the rationale presented in Section 2. In this scenario, a double strict liability rule, where both injurers are subject to strict liability but only one pays the damage (fine) to the state, may be effective if the victim's incentive to exercise care can be disciplined. Under the double strict liability rule, the victim lacks the incentive to exercise care, as full compensation is assured, regardless of their care level. However, if a negligent victim cannot be compensated for losses, they would exercise due care under this double strict liability rule, incorporating the defense of contributory negligence. If both injurers bear the full damage under the double strict liability rule, the victim, internalizing all accident risk, exercises due care. Consequently, social optimum can be achieved without the defense of contributory negligence.

De Chiara et al. (2021) provide a similar intuition, asserting that the strict liability rule outperforms the negligence rule, mitigating underinvestment. Note, however, that the role of investment differs in their sequential model compared with our simultaneous model. In a sequential model, investment does not directly impact accident probability but only indirectly influences it by increasing the manufacturer's care level. By contrast, investment directly affects accident probability in our simultaneous model.

5.2. Purchasing Decision

In this paper, we do not consider the purchasing decisions of human operators and the pricing decisions of the manufacturer, which constitute the core of the analysis in De Chiara et al. (2021).[21] However, we can conjecture that the market price is determined at a level equal to the manufacturer's marginal cost, comprising the production cost and care cost in a perfectly competitive market. This is because, under the negligence rule, the manufacturer does not bear the accident cost in equilibrium. If the manufacturer operates as a monopolist, incorporating the pricing decision into our model becomes nontrivial. This complexity arises from the dependence of the analysis on whether the manufacturer's observable pricing decision occurs before or after its unobservable decision to take care.

Exploring under what conditions or liability rules the competitive market price can substitute for the role of a liability rule in market-type torts is crucial. Intuitively, if the manufacturer takes minimal care, resulting in a high accident probability, market demand and, consequently, the price is low. This may incentivize the manufacturer to exercise sufficient care even without a liability rule. Thus, a natural question arises: Can the price alone internalize all externalities related to product-related accidents in the competitive market? Unfortunately, this appears not to be

[21] Several studies have delved into optimal liability rules within a market setting. An incomplete list encompasses works by Veljanovski (1979), Weingast, McBride, and Conant (1979), Shavell (1980), Polinsky (1980), Polinsky and Rogerson (1983), and Landes and Posner (1985).

the case. Landes and Posner (1985) demonstrate that high bargaining costs between the manufacturer and customers render care levels inefficient in the competitive market, irrespective of the liability rule. This contrasts with the Coase Theorem's possibility of efficiency when bargaining costs are low. The primary reason for this inefficiency is the negligible impact of individual agents' care choices on the competitive market price. However, under low transaction costs facilitating bargaining for contractual terms or in a monopolistic market, it becomes possible for the monopolist to have an incentive to take care. In this scenario, reducing the care level increases accident probability and decreases the price.

The insight of Landes and Posner (1985) remains applicable if consumers experience the same harm from product-related accidents and accurately perceive risk. Polinsky and Rogerson (1983) argue that if consumers underestimate risk, the negligence rule is preferable to the strict liability rule in a monopoly market. This is because it can counterbalance the monopolist's tendency to produce too little by shifting liability to consumers. Miceli et al. (2015) further demonstrate that if consumers suffer the same but varying harms or if they misperceive risk, negligence and no liability are efficient, but strict liability is not.

An intriguing extension would be to examine the optimal liability rule for accidents caused by fully autonomous vehicles in a perfectly competitive market. However, this is not straightforward because the buyer of an autonomous vehicle is no longer a plaintiff but another defendant. In this three-person game involving a seller, a buyer, and a victim, exploring how the results of Landes and Posner (1985) can be affected becomes a worthwhile endeavor.

VI. Concluding Remarks

In this paper, we present a straightforward model of accidents involving fully autonomous vehicles and proposed optimal liability rules to achieve socially efficient outcomes. We demonstrate that the social optimum for autonomous vehicles of level 4 or above can be realized through an HRL rule that assigns residual liability to human operators in the case of unilateral activity accidents and a VRL rule that assigns residual liability to victims in the case of bilateral activity accidents. This contrasts with the findings of Guerra et al. (2021), which are specific to semi-autonomous vehicles of level 3 or below.

Acknowledging the inherent simplification of our model, we recognize its limitations in capturing all the nuanced details of the real world. Nevertheless, we maintain that our analysis offers a clear insight into the optimal liability rules for accidents involving fully autonomous vehicles. We anticipate that this groundwork can serve as a valuable foundation for enriched extended models that incorporate more realistic features.

Appendix

Proof of Proposition 1: Let $a^S(\theta_H)$ and $x^S(\theta_H)$ be the solutions of (8) and (9). We first establish the validity of part (ii).

(ii) Differentiating (8) with respect to θ_H yields

$$\frac{da^S(\theta_H)}{d\theta_H} = \frac{p_a L}{\phi_{aa}} < 0, \tag{22}$$

and differentiating (9) with respect to θ_M yields

$$\frac{dx^S(a)}{d\theta_M} = -\frac{p_x L}{\psi_{xx}} < 0. \tag{23}$$

Subsequently, if $\theta_H, \theta_M \in (0,1)$, then $a^S(\theta_H) > a^S(1) = a^*$ and $x^S(\theta_M) > x^S(1) = x^*$. If $\theta_H = 1$, $\theta_M = 0$; therefore, $a^S(1) = a^*$ but $x^S(0) > x^*$. If $\theta_H = 0$, $\theta_M = 1$; hence, $a^S(0) > a^*$ but $x^S(1) = x^*$. This completes the proof of (ii).

(i) Suppose $a^S = a^*$ and $x^S = x^*$. Equations (8) and (9) imply that

$$B'_H(a^*) - p_a(a^*)\theta_H L = 0,$$
$$c'_M(x^*) + p_x(x^*)\theta_M L = 0.$$

If $\theta_H = 1$, then $\theta_M = 0$, so (9) is violated. If $\theta_H < 1$, (8) is not satisfied, implying a contradiction. Therefore, that $a^S = a^*$ and $x^S = x^*$ are impossible. □

Proof of Proposition 2: $a = a^*$ is the best response of D_H to $x = x^*$. Now, we consider the decision of D_M. Under the negligence rule, the cost function of D_M is given by

$$C_M^N(a,x) = \begin{cases} c_M(x) + p(a,x)\theta_M L & \text{if } x > \overline{x} \\ c_M(x) & \text{if } x \le \overline{x}. \end{cases} \tag{24}$$

If $\theta_M = 1$, the best response of D_M to $a = a^*$ is $x = \overline{x} = x^*$. D_M has no reason to choose $x < \overline{x}$ because $c'_M(x) < 0$. If $x > \overline{x}$, we have $\frac{dC_M^N(a^*,x)}{dx} = \frac{dC_M^S(a^*,x)}{dx} > 0$ because $x = x^* = \overline{x}$ minimizes $C_M^S(a^*,x)$, implying that $x^N = x^*$.

Now, it remains to show that $x^N = x^*$ for any $\theta_M (\ge \underline{\theta}_M)$ for some $\underline{\theta}_M \in (0,1)$. Note that

$$C_M^N(a^*,x^*) < \lim_{x \to x^*+0} C_M^N(a^*,x), \tag{25}$$

because $p(a^*,x)\theta_M L > 0$ for any $\theta_M > 0$. $x^S(\theta_M)$ is continuously decreasing as θ_M decreases, so $\min_x C_M^S(a^*,x^S(\theta_M);\theta_M)$ is also continuous with respect to θ_M. Considering that $C_M^N(a^*,x) = C_M^S(a^*,x)$ for any $x > \overline{x}$ by continuity of C_M^S with respect to θ_M, then $C_M^N(a^*,x^*) < \lim_{x \to x^*+0} C_M^N(a^*,x)$ for any $\theta_M \geq \underline{\theta}_M = 1-\varepsilon$ for some $\varepsilon > 0$. □

Proof of Proposition 3: By abusing notation, let x^N and a^N be the equilibrium care level of D_M and the equilibrium activity level of D_H under this negligence rule. Assume that $\overline{x} = x^N$. Under this rule, the manufacturer's cost function is

$$C_M^N(a,x) = \begin{cases} c_M(x) + p(a,x)L & \text{if } x > \overline{x} \\ c_M(x) & \text{if } x \leq \overline{x}. \end{cases}$$

Clearly, $x^N \not< \overline{x}$ because $c_M'(x) < 0$, and $x^N \not> \overline{x}$ because $\overline{x} = \arg\min_x c_M(x) + p(a,x)L$. As $C_M^N(a,\overline{x}) < c_M(\overline{x}) + p(a,\overline{x})L$, this implies that $x^N = \overline{x} = x^*$.

In addition, the human operator chooses a to maximize

$$\phi = B_H(a) - C_H^N(a,\overline{x}) = B_H(a) - p(a,\overline{x})L.$$

The first-order condition requires

$$\phi_a = B_H'(a^N) - p_a(a^N)L = 0. \tag{26}$$

This coincides with (4), implying that $a^N = a^*$. □

Proof of Proposition 4: Given that D_H chooses $a = a^*$, assume that $x = \overline{x}$. Then, the victim's cost function can be written as

$$C_P(a,x,y) = \begin{cases} c_P(y) + p(a^*,x^*,y)L & \text{if } y > \overline{y} \\ c_P(y) & \text{if } y \leq \overline{y}. \end{cases} \tag{27}$$

Given that $x = \overline{x}$, P will never choose $y > \overline{y}$ because $p(a^*,x^*,y)L > 0$ implies that $c_P(y) < c_P(y) + p(a^*,x^*,y)L$ for any $y > \overline{y}$. In addition, P will never choose $y < \overline{y}$ because $c_P'(y) < 0$, meaning that choosing $y = \overline{y} = y^*$ is optimal.

Now, we consider the best response of D_M to $y = \overline{y}$. The manufacturer's cost function is

$$C_M(a,x,y)=\begin{cases} c_M(x)+p(a^*,x,y^*)\theta_M L & \text{if } x>\overline{x} \\ c_M(x) & \text{if } x\le\overline{x}. \end{cases} \tag{28}$$

As is shown in Proposition 2, by the continuity argument, $\underline{\theta}_M\in(0,1)$ such that for all $\theta_M\ge\underline{\theta}_M$, $x=\overline{x}=x^*$.

Finally, if $x=\overline{x}$ and $y=\overline{y}$, D_H is fully liable for the whole damage under our hybrid liability rule. Taking the efficient activity level, $a=a^*$, is optimal for them. □

Proof of Proposition 5: The victim's problem is the same as in the case of Proposition 4, so choosing $x=\overline{x}$ remains optimal for the manufacturer. Under this rule where D_M is the primary liability bearer, the liability share of D_M is $\theta_M=1$ for any $x>\overline{x}$. This is a special case of $\theta_M>\underline{\theta}_M(>0)$ in Proposition 4, thus completing the proof. □

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, 127–171.

Ben-Shahar, O. (2016), "Should Carmakers Be Liable When a Self-Driving Car Crashes?" Forbes, September 22. https://www.forbes.com/sites/omribenshahar/2016/09/22/should-carmakers-be-liable-when-a-self-driving-car-crashes/2/#309606f74378.

Calabresi, G. (1961), "Some Thoughts on Risk Distribution and the Law of Torts," *Yale Law Journal*, 70, 499–553.

__________ (1970), *The Costs of Accidents*, Yale University Press, New Haven.

De Chiara, A., I. Elizaldeb, E. Manna and A. Segura-Moreiras (2021), "Car Accidents in the Age of Robots," *International Review of Law and Economics*, 68, 1–14.

Edlin, A. (1998), "Per-Mile Premiums for Auto Insurance," NBER Working Paper No. w6934.

Green, J. (1976), "On the Optimal Structure of Liability Law," *Bell Journal of Economics*, 553-574.

Guerra, A., F. Parisi, F., and D. Pi (2021), "Liability for Robots II: An Economic Analysis," *Journal of Institutional Economics*, 1–16.

Kornhauser, L. A., and R. Revesz (1989), "Sharing Damages Among Multiple Tortfeasors," *Yale Law Journal*, 98, 831–884.

Landes, W., M., and R. A. Posner (1980), "Joint and Multiple Tortfeasors: An Economic Analysis," *Journal of Legal Studies*, 9, 517–555.

Landes, W., and R. Posner (1985), "A Positive Economic Theory of Products Liability," *Journal of Legal Studies*, 14, 535–567

Logue, K. D. (2019), "The Deterrence Case for Comprehensive Automaker Enterprise Liability," *Journal of Law and Mobility*, 2019, 1–31.

Miceli, T., K. Segerson, and S. Wang (2015), "Products Liability when Consumers Vary in their Susceptibility to Harm and May Misperceive Risk," *Contemporary Economic Policy*, 33, 468-476.

Polinsky, M. (1980), "Strict Liability vs. Negligence in a Market Setting," *American Economic Review*, 70, 363–367.

Polinsky, M., and W. Rogerson (1983), "Products Liability, Consumer Misperceptions, and Market Power," *Bell Journal of Economics*, 14, 581–589.

SAE (2021), "SAE J3016 Levels of Driving Automation," available at https://www.sae.org/blog/sae-j3016-update.

Schwab, K. (2015), "The Fourth Industrial Revolution: What it Means and How to Respond," *Foreign Affairs*, December 12.

Shavell, S. (1980), "Strict Liability versus Negligence," *Journal of Legal Studies*, 9, 1–25.

________ (2020), "On the Redesign of Accident Liability for the World of Autonomous Vehicles," *Journal of Legal Studies*, 49, 243–285.

Talley, E. (2019), "Automatorts: How Should Accident Law Adopt to Autonomous

Vehicles," *Lessons from Law and Economics*, pp. 3.

Veljanovski, C. G. (1979), "'Economic' Myths about Common Law Realities – Economic Efficiency and the Law of Torts," Working paper no. 5, Centre Socio-Legal Stud., Oxford University.

Vickrey, W. (1968), "Automobile Accidents, Tort Law, Externalities and Insurance: An Economist's Critique," *Law and Contemporary Problems*, 3, 464.

Vladeck, D. C. (2014), "Machines without Principals: Liability Rules and Artificial Intelligence," *Washington Law Review*, 89, 117–150.

Weingast, B. R., M. E. McBride, and J. L. Conant (1979), "Product Safety and Consumer Information: The Impact of Liability Assignment and Standards Regulation," Working paper no. 45, Center Stud. Amer. Bus., Washington University (St. Louis).

인공지능의 법경제학: 완전자율주행차 사고에 대한 최적 책임원칙*

김 정 유**

초 록 본 연구에서는 완전자율주행차 사고에 대한 최적 책임원칙을 고려한다. 일방의 활동수준만 사고확률에 영향을 주는 사고에서는 법정 활동수준을 집행하기 어려운 경우, 활동수준을 결정하는 인간작동자에게는 엄격책임원칙을, 과실 정도를 결정하는 자동차 제조업자와 피해자에게는 비교과실 또는 비례과실책임을 적용하는 것이 사회적으로 효율적임을 입증한다. 하지만 법정 활동수준을 집행할 수 있는 경우에는 정반대의 책임원칙이 효율적이다. 이 경우 인간작동자는 법정 활동수준을 준수하고 제조업자는 법정주의수준을 준수하여 책임을 면하고, 피해자는 잔여책임을 부담함으로써 역시 효율적인 주의를 기울 이게 된다. 이러한 책임원칙은 쌍방의 활동수준이 모두 사고확률에 영향을 주는 경우에도 최적의 사회적 효율성을 보장한다.

핵심 주제어: 인공지능, 자율주행차, 부진정연대책임, 책임원칙, 잔여책임
경제학문헌목록 주제분류: K13

투고 일자: 2022. 11. 25. 심사 및 수정 일자: 2023. 1. 30. 게재 확정 일자: 2023. 3. 15.
* 이 논문은 2022년 대한민국교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2022S1A5A2A0304932311).
** 경희대학교 경제학과 교수, e-mail: jyookim@khu.ac.kr