Canadian Journal of Law & Jurisprudence XXXIV No.1 February 2021, 105-147

© The Author(s) 2020 doi: 10.1017/cjlj.2020.3

Tort Law, Corrective Justice and the Problem of Autonomous-Machine-Caused Harm

Pinchas Huberman

I. Introduction

Developments in artificial intelligence and robotics promise increased interaction between humans and autonomous machines, presenting novel risks of accidental harm to individuals and property.[1] This essay situates the problem of autonomous-machine-caused harm within the doctrinal and theoretical framework of tort law, conceived of as a practice of corrective justice. The possibility of autonomous-machine-caused harm generates fresh doctrinal and theoretical issues for assigning tort liability. Due to machine-learning capabilities, harmful effects of autonomous machines may be untraceable to tortious actions of designers, manufacturers or users.[2] As a result, traditional tort doctrine—framed by conditions of foreseeability and proximate causation—would not ground liability.[3] Without recourse to compensation, faultless victims bear the accident costs of autonomous machines. This doctrinal outcome reflects possible incompatibility between tort's theoretical structure of corrective justice and accidents involving autonomous machines. As a practice of corrective justice, tort liability draws a normative link between particular defendants and plaintiffs, as doers and sufferers of the same tortious harm, grounding defendants' agent-specific obligations to repair the harm. Where accidents are caused by autonomous machines, the argument goes, the essential link between defendants and plaintiffs is severed; since resulting harm is not legally attributable to the human agency of designers, manufacturers or users, victims have no remedy in tort.

The issue of liability for autonomous-machine-caused harm implicates a cluster of policy considerations: Which standard of liability—e.g. negligence or strict liability—best incentivizes cost-efficient and safe technological innovation?

I am grateful to Professors Peter Benson and Bruce Chapman for valuable feedback and comments on earlier versions of this essay, which constituted part of my LLM thesis at UofT. I am also grateful to Ben Ohavi and Yona Gal for helpful and lively discussions about ideas in this essay.

Pinchas Huberman, incoming LLM candidate fall 2020, Yale Law School.
pinchas.huberman@yale.edu

1. Ryan Calo, "Robotics and the Lessons of Cyberlaw" (2015) 103 Cal L Rev 513 at 532-34.
2. David C Vladeck, "Machines Without Principals: Liability Rules and Artificial Intelligence" (2014) 89 Wash L Rev 117 at 121-23.
3. Calo, *supra* note 1 at 542; Peter M Asaro, "The Liability Problem for Autonomous Artificial Agents" in *Ethical and Moral Considerations in Non-Human Agents, 2016 AAAI Spring Symposium Series* (AAAI Press, 2016) 190 at 191 [Asaro, "Liability Problem"]; Curtis EA Karnow: "The Application of Traditional Tort Theory to Embodied Machine Intelligence" in Ryan Calo, A Michael Froomkin & Ian Kerr, eds, *Robot Law* (Edward Elgar, 2016) 51 at 63-74.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

Does tort law fairly distribute accident costs of autonomous machines in proportion to their benefits? Would alternative liability regimes provide victims with more-satisfactory access to compensation for accidental harm caused by autonomous machines? Admittedly, this essay does not meaningfully address these broader issues of law and policy. Its primary focus is liability analyses under the common law of tort, interpreted through the normative lens of corrective justice. In this view, tort law interprets individuals' interpersonal rights and duties with respect to designing, manufacturing and using autonomous machines. Tort law evaluates what individuals legally owe each other when setting into motion functionally independent machines, deliberately deployed to interact with other individuals and property. In this respect, successful tort analyses need to conceptualize the normative significance of autonomous machines within human social relations. While tort law does not need to take—nor should it take—a comprehensive position on the possibility of artificial intelligence, it needs to suitably assess the normative impact of autonomous machines on interpersonal rights and duties of legal subjects.

To some extent, courts will not encounter automated machines for the first time; they have been used in both industrial and recreational settings for decades, and courts have already faced robot-specific legal issues. Notably, judicial pronouncements reveal an attitude representing robots as lifeless entities, stripped of agency and totally lacking in autonomy, discretion and spontaneity.[4] However, this denigrating view of automated machinery needs to be reconsidered in light of current technological advances.[5] Recent work in robotics, artificial intelligence and computer technologies has advanced in two general ways.[6] First, computer systems can learn from experience through analyzing huge amounts of instructive data at extraordinarily high speed.[7] Second, with a combination of sensors and actuators, embodied computer systems can interact with their external environments, enabling dynamic physical tasks.[8] These advances are triggering deployment of sophisticated technologies in numerous new roles, projected to be

4. Ryan Calo, "Robots as Legal Metaphors" (2016) 30:1 Harv JL & Tech 209 at 221-25. Calo notes several examples. In one illuminating instance, a Maryland court considered whether an animatronic robot's dancing and singing constituted a "performance" within the meaning of a statute imposing tax on food served in restaurants "where there is furnished a performance." It was held that robots do not "perform" (within the meaning of the statute) since they lack skill and potential for spontaneous imperfection. See *Comptroller of the Treasury v Family Entertainment Centers*, 519 A (2d) 1337 (Md Ct Spec App 1987) at 1338. Moreover, some courts have used the term 'robot' metaphorically to justify decisions relieving defendants' responsibility for tortious harm. For instance, individuals are compared to robots where they are seen to act without agency—as mere instruments of others—and cannot be held legally responsible in tort. See *Frye v Baskin*, 231 SW (2d) 630 (Mo Ct App 1950) at 635. Likewise, pursuant to the "alter ego" theory, a corporate veil can be pierced by showing that controlled corporations acted "robot-like" in "mechanical response" to a controller's "pressure on its buttons". See *Culbreth v Amosa (Pty) Ltd*, 898 F (2d) 13 (3d Cir 1990) at 15.

5. Calo, "Robots as Legal Metaphors", *supra* note 4 at 226.

6. Jerry Kaplan, *Humans Need Not Apply: A Guide to Wealth and Work in the Age of Artificial Intelligence* (Yale University Press, 2015) at 4.

7. *Ibid*.

8. *Ibid* at 5.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

significant parts of our social fabric in coming years. For instance, we anticipate robots to perform surgeries,[9] automated drones to deliver packages[10] and automated vehicles to drive without human assistance.[11] Robots are also expected to serve as shopping assistants,[12] therapeutic aids and pets,[13] and as caregivers for the elderly.[14] Moreover, at the present time, security robots monitor shopping malls,[15] warehouse robots independently organize and fetch storage items to fill orders,[16] and robotic vacuum-cleaners operate autonomously in cluttered environments, scooting around obstacles without human involvement.[17] Autonomous machines are generally expected to be beneficial. For example, driverless cars promise to reduce accidents, increase transportation efficiency, enhance individuals' mobility, and reduce emissions.[18] However, as noted, autonomous machines also pose new problems, such as risks of autonomous-machine-caused harm to individuals and property, without clear legal resolution. There have already been reported accidents involving self-driving cars and warehouse robots, and a handful of legal suits for personal injury and wrongful death.[19]

---

9. Meera Senthilingham, "Would You Let a Robot Perform Your Surgery By Itself?", *CNN* (12 May 2016), online: https://www.cnn.com/2016/05/12/health/robot-surgeon-bowel-operation/index.html.

10. Laura Stevens & Georgia Wells, "UPS Uses Drone to Deliver Packages to Boston-Area Island", *Wall Street Journal* (23 September 2016), online: https://www.wsj.com/articles/ups-uses-drone-to-deliver-package-to-boston-area-island-1474662123.

11. Heather Somerville, "Uber, Transitioning to Fleet Operator, Orders 24,000 Driverless Cars from Volvo", *Insurance Journal* (21 November 2017), online: https://www.insurancejournal.com/news/national/2017/11/21/471938.htm.

12. Jana Kasperkevic, "Swiss Police Release Robot that Bought Ecstasy Online", *The Guardian* (22 April 2015), online: https://www.theguardian.com/world/2015/apr/22/swiss-police-release-robot-random-darknet-shopper-ecstasy-deep-web.

13. Jacqueline Howard, "Robot Pets Offer Real Comfort", *CNN* (1 November 2017), online: https://www.cnn.com/2016/10/03/health/robot-pets-loneliness/index.html.

14. Adriana Barton, "Cylons, They Are Not. These Intelligent and Friendly Robots Are Designed To Help the Elderly Live A Better Life", *The Globe and Mail* (26 August 2018), online: https://www.theglobeandmail.com/life/health-and-fitness/article-cylons-they-are-not-these-intelligent-and-friendly-robots-are/.

15. James Vincent, "Mall Security Bot Knocks Down Toddler, Breaks Asimov's First Law of Robotics", *The Verge* (13 July 2016), online: https://www.theverge.com/2016/7/13/12170640/mall-security-robot-k5-knocks-down-toddler.

16. Alistair Barr, "Amazon.com to Buy Kiva Systems for $775 Million", *Reuters* (19 March 2012), online: https://www.reuters.com/article/us-amazoncom/amazon-com-to-buy-kiva-systems-for-775-million-idUSBRE82I11720120319.

17. John Markoff, "What Comes After the Roomba", *The New York Times* (21 October 2018), online: https://www.nytimes.com/2018/10/21/business/what-comes-after-the-roomba.html?rref=collection%2Ftimestopic%2FiRobot%20Corporation&action=click&contentCollection=business&region=stream&module=stream_unit&version=latest&contentPlacement=1&pgtype=collection

18. Ryan Abbott, "The Reasonable Computer: Disrupting the Paradigm of Tort Liability" (2018) 86:1 Geo Wash L Rev 1 at 42.

19. Daisuke Wakabayashi, "Self-Driving Uber Car Kills Pedestrian in Arizona Where Robots Roam", *The New York Times* (19 March 2018), online: https://www.nytimes.com/2018/03/19/technology/uber-driverless-fatality.html; Peter Holley, "After Crash Injured Motorcyclist Accuses Robot-Driven Vehicle of 'Negligent Driving'", *Washington Post* (25 January 2018), online: https://www.washingtonpost.com/news/innovations/wp/2018/01/25/after-crash-injured-motorcyclist-accuses-robot-driven-vehicle-of-negligent-driving/?utm_term=.ca85154c515e; The Fernandez Firm, "What Happens When a Robot Causes Wrongful Death",

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

This essay proceeds as follows: Part II reviews the salient features of autonomous machines that threaten to disrupt the structure of interpersonal rights and duties in tort.[20] It is not a comprehensive technical or philosophical account of artificially intelligent mechanisms. Rather, it is an account of autonomous machines' features that are especially pertinent to the application of tort law. The aim is to conceptualize the normative position of autonomous machines within tort relations—i.e. their effect on interpersonal rights and duties—to guide the application of tort doctrine. For this reason, in the remainder of this essay, I refer to autonomous machines as autonomous agents (AAs), not as robots or artificial intelligences. My intent is to signify the legal impact of autonomous machines, not their technical or philosophical connotations.

Part III describes the doctrinal and theoretical structure of tort law, as embodying a practice of corrective justice. Arguably, the doctrinal challenge of discovering a basis of tort liability for AA-caused harm relates to tort's theoretical structure of corrective justice, which resists straightforward application to AA-mediated social relations. To effectively grapple with the doctrinal challenge, tort's theoretical structure must be adequately addressed.

Part IV examines several doctrinal strategies to ground tort liability for AA-caused harm under the common law of negligence and strict liability. It concludes that under these laws, AA-caused losses likely lie where they fall. It argues that this doctrinal outcome is unsatisfactory, however, even from the viewpoint of corrective justice, as it misunderstands the mediating role of AAs within human relations. AAs are purposely supplied and utilized by their manufacturers and users knowing that they entail uncontrollable possibility of harm, at least in isolated cases. Since human deployers leverage AAs' functional independence to advance their purposes, they should assume some responsibility for AA-caused harm. Straightforward application of negligence and strict liability doctrines is inadequate, and alternative liability standards and regimes should be considered.

Part V surveys the emerging academic literature about tort liability for AA-caused harm, revealing three strands of thought: to affirm the applicability of tort, modify it, or abandon it altogether for alternative liability regimes. It suggests that it may be necessary to reconceive tort principles and AAs' legal classification to suitably assimilate AAs within the tort framework. In advancing the common law of tort, analogical reasoning is essential. AAs should be compared to, and distinguished from, familiar tort categories. This demands rigorous thinking about AAs' salient features that should have normative impact on rights and duties of legal persons. The essay concludes by noting a possible legal conception of AAs: as legal agents without legal personhood. Under this approach, human

Medium (30 March 2017), online: https://medium.com/@thefernandezfirm/what-happens-when-a-robot-causes-wrongful-death-3d1f4f7e9711; Seth Baum & Trevor White, "When Robots Kill", *The Guardian* (23 July 2015), online: https://www.theguardian.com/science/political-science/2015/jul/23/when-robots-kill.

20. See Jack B Balkin, "The Path of Robotics Law" (2015) 6 Cal L Rev Circuit 45 at 47-48.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

(or corporate) deployers could be vicariously liable for the tortious acts of AAs committed in the course of deployment.

II. Autonomous Agents: Embodiment, Emergence and Machine Learning

As a general matter, AAs are machines capable of performing complex tasks previously requiring human intelligence, such as abstract problem-solving, object recognition, natural language processing and social interaction.[21] AAs have abilities to sense phenomena, process what they sense, and act upon the world.[22] In distinguishing AAs from other technologies in their threat to disrupt tort law, two qualities should be emphasized: their capacities for embodiment and emergence.[23] First, AAs are embodied in the sense that they take some corporeal form, organized to physically act upon the world.[24] AAs' embodiment enables direct interaction with, and impact on, individuals and property. Second, AAs are emergent in that their behaviours are not entirely pre-programmed in their own blocks of code; instead, AAs adapt to circumstances and learn from experience, updating their algorithms through machine learning techniques.[25] Programmers specify goals for AAs to achieve, permitting AAs to train on illustrative examples of their tasks, and to author their algorithms through processes of trial-and-error.

Machine learning is a developmental process where AAs improve particular behavioural abilities through repeated exposure to "information-rich environments."[26] An increasingly important machine learning method is artificial neural network learning. Neural networks have sets of input and output nodes representing features of input data and desired output action, as well as layers of hidden nodes that process the input data, extracting information relevant to desired outputs.[27] As additional input data is received, neural networks are trained by learning algorithms which adjust numerically weighted connections between the layers of nodes to optimize the relationship between its input and output.[28] Through repeated "training runs with input datasets," the machine's output—its "behaviour"—is enhanced for its particular task.[29]

Machine learning techniques do not necessarily replicate human or animal cognitive processes; AAs 'learn' in a functional sense, entailing the capacity to change actions and improve performance through training on illustrative

21. Shannon Vallor & George Bekey, "Artificial Intelligence and the Ethics of Self-Learning Robots" in Patrick Lin, Keith Abney & Ryan Jenkins, eds, *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence* (Oxford University Press, 2017) 338.
22. Calo, "Robotics and the Lessons of Cyberlaw", *supra* note 1 at 529. This is neatly phrased as the sense-think-act paradigm.
23. *Ibid* at 532.
24. *Ibid* at 534.
25. *Ibid* at 538-39.
26. Vallor & Bekey, *supra* note 21 at 340.
27. *Ibid*.
28. *Ibid*.
29. *Ibid*.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

datasets.[30] Machine learning algorithms are often statistical, detecting patterns in data that enable the algorithms to automatically build—without additional manual programing by human engineers—internal computer models of a phenomenon to make further predictions or automated decisions about future data it receives, i.e. new examples of the phenomenon at issue.[31] AAs' abilities to detect relevant patterns and build useful internal models—and their resulting accuracy in achieving assigned tasks—improves as the algorithm gathers and senses more examples of the phenomenon; in this sense, additional data facilitates refinement of the algorithm, improving overall performance.[32]

Emergent AAs utilizing machine learning algorithms operate with partial independence from their programmers. They do not simply contain manually programed sets of pre-determined rules, but models automatically and incrementally formed and refined through processes of pattern detection in new data sets.[33] Computer systems may be built initially with under-specified code—instead relying on high-level goals—where precise solutions are unknown in advance and can benefit from experimentation in real environments.[34] For instance, to design a robotic arm that works in dynamic environments, the arm's movements cannot be pre-determined by a precisely specified algorithm due to the many environmental contingencies it will encounter; instead, machine learning techniques may be applied to derive a model for the arm's movement.[35] While the model may start with largely unknown parameters, the learning algorithm modifies the parameters as it encounters new conditions to achieve optimal performance of its objective.[36] Consequently, while the learning algorithm and data are initially supplied by human programmers, the ultimate algorithm controlling the robotic arm is "discovered from data" by the learning algorithm.[37] Moreover, through reinforcement learning, systems learn in their final operating environments, optimizing their parameters through processes of trial and-error.[38] In this sense, the possibility of error is required for learning and emergent behaviour; it is a feature of learning systems, not a product flaw.[39]

Emergent capabilities exist on a spectrum. Increased emergence is required to the extent embodied AAs need to rearrange their code in real time in reaction to higher degrees of environmental constraints.[40] AAs interact, and form larger systems, with their environments: while environmental inputs shape AAs'

30. Harry Surden, "Machine Learning and the Law" (2014) 89 Wash L Rev 87 at 89.
31. *Ibid* at 91-95.
32. *Ibid* at 93.
33. *Ibid* at 94.
34. See Deven R Desai & Joshua A Kroll, "Trust but Verify: A Guide to Algorithms and the Law" (2017) 31:1 Harv JL & Tech 1 at 26.
35. *Ibid* at 27.
36. *Ibid*.
37. *Ibid* at 28.
38. Andreas Matthias, "The Responsibility Gap: Ascribing responsibility for the actions of learning automata" (2004) 6 Ethics and Information Technology 175 at 179.
39. *Ibid*.
40. See Karnow, *supra* note 3 at 56-60.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

activities, AAs' outputs, in turn, alter their environments.[41] As a result, larger AA-environmental systems become progressively more unpredictable than they were initially.[42]

Accordingly, AAs' relative independence extends beyond not being controlled by humans in real time, but to their self-selection of methods to achieve programmed goals.[43] This is not to say that AAs have discretionary powers or wills of their own. They are deterministic systems, their outputs defined by inputs received.[44] Moreover, AAs' ultimate goals are those of their human designers and users, not their own. Nevertheless, due to AAs' emergence, it is difficult—if not impossible—to fully trace or predict connections between environmental inputs and ensuing changes in AAs' algorithms and behaviours. Furthermore, the behaviour of neural networks can only be evaluated by applying test patterns and then observing its output; it is not possible to simply look at information stored or represented in the network.[45] AAs' activities are then characteristically unpredictable. Ryan Calo describes emergent AAs as "unpredictable by design."[46] Multiple design teams or companies may contribute to software and hardware components that comprise a final product design, and no single group can fully know how the system will react to new inputs.[47] This is particularly true when machines continue to learn in their final operating environments. Since programmers do not have control over these environments—including the various people and situations machines will encounter—they cannot maintain full control over their product.[48] In these instances, users and programmers can expect AAs to make errors on the way to optimizing output.[49] Even well-trained and reliable networks or systems can fail to model desired outputs while acting in the real world, leading to AA-caused harm.[50] While such harm is reflective of misalignment between code, network weightings and programmers' goals, this misalignment need not be introduced by, or knowable to, programmers themselves.[51] Crucially, where AAs have combined capacities of emergence and embodiment, their principled unpredictability is coupled with potential to cause physical harm to individuals and property.

AAs' activities carry distinctive normative implications. Individuals are at risk of being directly impacted—negatively or otherwise—by AAs in ways unforeseeable to their human designers or users, whether in principle or practice.[52]

---

41. *Ibid*.
42. *Ibid*.
43. *Ibid*.
44. Neil M Richards & William D Smart, "How Should the Law Think About Robots" in Ryan Calo, A Michael Froomkin & Ian Kerr, eds, *Robot Law* (Edward Elgar, 2016) 3 at 18.
45. Matthias, *supra* note 38 at 178.
46. Calo, "Robotics and Lessons of Cyberlaw", *supra* note 1 at 542.
47. Wendell Wallach & Colin Allen, "Moral Machines: Teaching Robots Right from Wrong" (Oxford University Press, 2009) at 39.
48. Matthias, *supra* note 38 at 182.
49. *Ibid*.
50. Vallor & Bekey, *supra* note 21 at 343.
51. *Ibid*.
52. Asaro, "Liability Problem", *supra* note 3 at 191: "some artificial agents may be unpredictable in principle, and many will be unpredictable in practice."

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

Where AAs utilize machine learning techniques to author their controlling algorithms, they attain some independence from the humans they serve. This independence has two dimensions: First, machine learning processes and engagement with novel circumstances undermines human ability to fully predict or foresee future AA-activities. In this negative sense, AAs' independence may be described as the negation of control and foreseeability on the part of their human designers and users. Second, through machine learning processes, controlling algorithms are effectively designed by AAs' learning algorithms. In this affirmative sense, AAs formulate and change behaviours through internally caused processes, usefully represented as a kind of functional agency or autonomy.

This analysis is not technical, but conceptual and normative: it is directed at isolating the legally relevant causal force of AAs' outward effects for liability purposes. Can AAs' outward effects be attributed to human designers or users though unforeseeable? Is it accurate to describe AAs as performing actions? Are AAs initiators—i.e. their own causes—of their actions? Does AAs' mediation of human interaction serve to link users or designers of AAs and their victims, or does AAs' mediation sever the ties of human interrelation? Since normative considerations of foreseeability, action and causation are central to determinations of interpersonal responsibility in tort law, it is crucial to clarify the normative position of AAs within human relations.

Recent academic commentary has begun this task. Ugo Pagallo emphasizes three qualities of AAs—interactivity, autonomy and adaptability—as significant considerations of AAs' agency, challenging assumptions about human (or corporate) legal responsibility for AA-caused harm. Pagallo notes that AAs interactively perceive and alter their own properties in response to environmental stimuli.[53] In so doing, AAs adapt to changes in circumstances by modifying and improving—without direct human intervention—the rules that change their inner states, thereby taking control of their actions, a kind of autonomy.[54] Likewise, Wolf Loh and Janina Loh acknowledge the conceptual possibility of emergent AAs having functional equivalents to rational and agential autonomy.[55] Rational autonomy, understood as "action on the basis of the weightiest of reasons," has a functional equivalent represented by algorithms and corresponding external data.[56] Agential autonomy, which involves identifying certain behaviours as "genuine actions," that is, "not entirely determined by external factors," has a functional equivalent represented by AAs' abilities to "change internal states without external stimuli."[57] Relatedly, Ryan Abbott focusses on the nature of control and independence, suggesting that AAs may be conceived

53. Ugo Pagallo, *The Laws of Robots: Crimes, Contracts and Torts* (Springer Netherlands, 2013) at 38.
54. *Ibid*.
55. Wolf Loh & Janina Loh, "Autonomy and Responsibility in Hybrid Systems" in Patrick Lin, Keith Abney & Ryan Jenkins, eds, *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence* (Oxford University Press, 2017) 36.
56. *Ibid* at 39.
57. *Ibid*.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

as entities that "control the *means* of completing tasks, regardless of their programming."[58] Finally, Samir Chopra and Laurence White suggest several ways to think about the possibility of AAs' autonomy: whether AAs' behaviour is determined by their own experiences; whether AAs initiate, not merely mediate, interactions; whether AAs operate without human control; whether AAs have social abilities to interact with humans or other AAs; whether AAs have proactive ability to initiate goal-directed activity; and whether AAs have reactive abilities to perceive and respond to changes in environment.[59] If several of these factors are combined, the case for AAs' autonomy is strengthened. Chopra and White also argue that the law may adopt an intentional stance toward AAs as intentional agents—i.e. practical reasoners engaged in goal-directed activities—potentially grounding AAs' status as legal agents or persons.[60]

The nature of AAs' functional autonomy and agency is central to assessing their normative position within social relations. However, for purposes of tort liability, AAs' functional autonomy should not be examined in isolation from tort's institutional procedures, doctrines and principles. I now turn to tort's doctrinal and theoretical structure, including its use of normative concepts like foreseeability, wrongdoing and causation. The broad issue is how tort law applies to instances of AA-caused harm, and focus is directed at how to construe AAs' impact on interpersonal rights and duties in tort.

III. Tort Liability, Corrective Justice and AA-Caused Harm

Tort law is a system of practical reason: it provides individuals with primary norms defining rightful conduct, as well as remedial norms empowering plaintiffs to bring suit, adduce evidence, and draw certain inferences that obligate defendants to repair their losses.[61] To this end, tort law sets out normative criteria for liability, including specific procedural structures and substantive doctrines. In particular, tort law assumes a bilateral relation as its subject: its doctrines determine whether a particular defendant is liable to repair a particular plaintiff's loss.[62] In the context of accidents causing personal injury or property damage, a set of principles—duty of care, breach of standard of care (tortious conduct), factual and proximate causation—determine whether there is a legally relevant link between a particular defendant's action and the plaintiff's loss, grounding

58. Abbott, *supra* note 18 at 5.
59. Samir Chopra & Laurence F White, *A Legal Theory for Autonomous Artificial Agents* (The University of Michigan Press, 2011) at 9-11.
60. *Ibid* at 11-18.
61. Jules Coleman, *The Practice of Principle: In Defence of a Pragmatist Approach to Legal Theory* (Oxford University Press, 2003) at 10 [Coleman, *Practice of Principle*].
62. As a general matter, the account of tort adopted in this essay—as a practice of corrective justice—follows work by Ernest J Weinrib, *The Idea of Private Law* (Oxford University Press, 2012) [Weinrib, *Private Law*]; Arthur Ripstein, *Private Wrongs* (Harvard University Press, 2016); Peter Benson, "Misfeasance as an Organizing Normative Idea in Private Law" (2010) 60 UTLJ 731; Coleman, *Practice of Principle*, *supra* note 61; and Stephen Perry, "Responsibility for Outcomes, Risk and the Law of Torts" in Gerald J Postema, ed, *Philosophy and the Law of Torts* (Cambridge University Press, 2001) 72.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

the defendant's agent-specific duty to repair.[63] Conduct is deemed tortious where the defendant's activity poses unreasonable foreseeable risk—i.e. real and substantial risk beyond levels of risk that are assumed in ordinary patterns of interaction—to the plaintiff (to whom a duty is owed).[64] Where the tortious risk materializes, causing loss to an object of the plaintiff's right, the defendant is liable to repair the loss. In this way, the law ascribes normative attribution of the plaintiff's loss to the defendant's wrongful conduct, constituted by its unreasonable foreseeable risk to the plaintiff (in breach of duty).[65]

Accordingly, central to tort liability is the combination of an unreasonably risky action and causation of harm. Of particular importance is the foreseeability of risk, which informs the boundaries of tort duties, the standard of care, and proximate causation. The existence of a duty of care owed by defendant to plaintiff establishes the right kind of relationship for tort liability. According to Lord Atkin's authoritative formulation, individuals owe duties to those "persons who are so closely and directly affected by my act that I ought reasonably to have them in contemplation as being so affected when I am directing my mind to the acts or omissions which are called in question."[66] Tort duties are relational: they are owed to particular plaintiffs with respect to particular foreseeable risks of potential harm. This was emphasized by Chief Judge Cardozo in *Palsgraf*: "the risk reasonably to be perceived defines the duty to be obeyed, and risk imports relation; it is risk to another or to others within the range of apprehension."[67] Duties are informed by the kinds of harm that are reasonably foreseeable: while defendants must modify their conduct with respect to reasonably foreseeable harm to plaintiffs, their duties do not extend to risks that are not within the "range of apprehension," even if the defendant already owes the plaintiff a duty.[68] The standard of care then establishes the kind of conduct required by tort duties; defendants' failures to meet the standard of care constitutes breach of duty, i.e. tortious conduct. Tortious conduct entails imposition of (unreasonable) risk by defendant on plaintiff exceeding the ordinary degree of risk that careful people impose on each other, as typically assumed in ordinary patterns of interaction.[69] The proximate causation inquiry then ensures that the plaintiff's harm is not only causally attributable to the defendant's tortious act, but is also the kind of harm that is within the scope of the risk that makes the defendant's act tortious.[70] It compares the kind of harm suffered by the plaintiff with the nature of the duty owed and the wrongful conduct that occurred, confirming that the plaintiff's harm is a realization, not merely the result, of the

63. Weinrib, *Private Law*, *supra* note 62 at 169-70.
64. Perry, *supra* note 62 at 111.
65. Ripstein, *supra* note 62 at 91.
66. *Donoghue v Stevenson*, [1932] AC 562 (HL) at 580.
67. *Palsgraf v Long Island Railroad Co*, 162 NE 99 (NYCA 1928) at 100.
68. *Ibid*.
69. Ripstein, *supra* note 62 at 102; Perry, *supra* note 62 at 111.
70. *Overseas Tankship (UK) v Morts Dock & Engineering (The Wagon Mound, No 1)*, [1961] AC 388 (PC); *South Australia Asset Management Corp v York Montague Ltd*, [1997] AC 191 (HL).

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

defendant's breach of duty to the plaintiff.[71] Importantly, the determination of proximate cause—i.e. whether the resulting harm falls within the scope of the tortious risk—turns on the particular description of the tortious risk (i.e. the foreseeable and unreasonable risk).[72] A general or broad description of tortious risk increases the chances of finding proximate cause, while a more specific or narrow description of risk reduces this prospect.[73] The general rule is that plaintiffs do not need to show that the specific chain of events causing harm was foreseeable, but that the resulting harm is a type of harm that was foreseeable and made the initial act wrongful.[74] A risk description's appropriate level of generality or specificity is certainly not self-applying; the assessment is fluid and fact-sensitive, requiring judgment, on a case by case basis, about which details are important, or insignificant, to the description of tortious risk.[75]

---

71. Ripstein, *supra* note 62 at 91.

72. Robert Keeton, *Legal Cause in the Law of Torts* (Ohio State University Press, 1963) at 49.

73. *Ibid*. The difference between negligence and strict liability relates to the breadth of the tortious risk description. Under strict liability, the tortious aspect is described in such a general manner so as to render an activity itself tortious because it inherently imposes a degree of risk exceeding that which is typically assumed in ordinary patterns of interaction. Strict liability applies to resulting harm that is within the scope of the activity's extraordinary risk, following the same proximate causation structure as negligence: strict liability applies if, and only if, the resulting harm is a kind that renders the activity exceedingly dangerous (i.e. tortious).

74. *Hughes v Lord Advocate*, [1963] AC 837 (HL) [*Hughes*]; *School Division of Assiniboine South No. 3 v Hoffer*, [1971] 21 DLR (3d) 608 (MB CA) [*Assiniboine*]; *Jolley v Sutton London Borough Council*, [2000] 3 All ER 409 (HL).

75. Clarence Morris, "Duty, Negligence and Causation" (1952) 101 U Pa L Rev 189 at 196-98. As an illustration, consider contrasting proximate cause analyses in two influential cases, *Hughes*, *supra* note 74 and *Doughty v Turner Manufacturing Co, Ltd*, [1964] 1 QB 518 (CA) [*Doughty*]. In *Hughes*, a defendant worker negligently left an open manhole unattended with warning lamps next to it, attracting children who took and dropped the lamps into the manhole triggering an explosion that caused the children to fall in and sustain burning injuries. According to the House of Lords, the plaintiff's burning injury was within the scope of the defendant's negligent act, as the risk of a burning injury was substantial and foreseeable, i.e., it was the kind of injury that made the defendant's act negligent. The Court acknowledged that the injury occurred through an unforeseeable sequence of events, an explosion, rather than a foreseeable occurrence, such as the children dropping and breaking the lamp while exploring the manhole, causing burning injuries through direct contact with the flame. Nevertheless, the chosen description of risk was quite general—simply the risk of burning—and not a more specific description, such as risk of burning caused by coming into direct contact with a broken lamp. The unforeseeable means of its occurrence did not undermine the foreseeability of the harm. By contrast, consider *Doughty*, where an employee suffered a burning injury from an explosion of a cauldron of molten liquid due to a cement cover falling in and disintegrating. The Court found that the defendant employer negligently failed to provide safeguards against the cement cover falling in the cauldron and causing burning injuries from the splashing of molten liquid. However, since the injury in this case was due, not to splashing, but to an explosion below surface—which was unforeseeable—the injury was found to fall outside the scope of the defendant's negligent act. In other words, the Court's description of the tortious risk was quite specific and narrow—risk of burning from the *splashing* of molten liquid caused by the cement cover falling in—rather than a more general description, such as risk of burning by molten liquid caused by the cement cover falling in. In *Doughty*, the unforeseeable sequence leading to the burning's occurrence rendered the harm unforeseeable, falling outside the scope of the tortious risk. This contrasts starkly with *Hughes*, where the unforeseeable sequence leading to the burning's occurrence did not alter the evaluation of foreseeability of its kind of harm. These cases illustrate both the importance and difficulty of describing tortious risk at the appropriate level of generality. How relevant are the particular details of the means of the occurrence of harm, and do they comprise part of the description of tortious risk?

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

Emergent AAs—entities that are "unpredictable by design"—pose a challenge to discovering tort liability, which is grounded in the foreseeability of real and substantial risk. If AAs' algorithmic developments and harmful outward effects are unforeseeable, in principle or practice, it will be difficult to establish tortious conduct (i.e. breach of duty) on the part of AAs' owners, users, designers or manufacturers.[76] Absent foreseeable and unreasonable risk of AA-caused harm, the act of deploying or distributing AAs, in any particular instance, is not deemed tortious. Moreover, it may be difficult to link AA-caused harm to responsible human conduct, as its proximate cause, where AAs' harmful effects result from inner-algorithmic processes, impacted by (unknown) environmental inputs.[77] It is possible that the resulting harm will be more precisely attributable to some environmental input, rather than a tortious aspect of the initial act, if any. This problem is not entirely insurmountable. Since proximate causation of harm ultimately turns on the particular description of tortious risk, one may be inclined to construe broadly a tortious risk in deploying AAs: a risk of algorithmic error and harm. The resulting AA-caused harm could then be viewed as the kind of harm that makes the initial deployment tortious—satisfying the requirement of proximate causation—though the specific manner in which the harm occurs (due to algorithmic processes) is unknown or unforeseeable.[78] Nonetheless, it must be emphasized that risks are not tortious unless they are real and substantial—exceeding those typically assumed in ordinary patterns of interaction—and it is a mistake to assume that deploying AAs is extraordinarily risky, as I argue below in section IV. If AAs' dangerous behaviours and harmful effects stem from their unpredictable self-learning algorithms—not from specific risky acts of programming or operating the AAs—they are arguably not materializations of tortious risk.[79] This is not to say that AA-caused harm can never be traceable

Arguably, *Doughty* and *Hughes* exemplify contrasting—though perhaps both plausible—approaches, highlighting an indeterminacy in the proximate cause inquiry. This question is critical in the context of AA-caused harm: to what extent do designers or deployers of AAs need to be able to foresee particular algorithmic developments that cause harm? Is the act of designing, manufacturing or deploying an AA with emergent capabilities itself a tortious risk, regardless of specific algorithmic developments causing harm in the circumstances? Or do unforeseeable—and unlikely—algorithmic developments causing harm in the circumstances render the harm unforeseeable, and therefore, unattributable to an initial tortious act?

76. Karnow, *supra* note 3 at 63-64, 72-74; Pagallo, *supra* note 53 at 117.

77. Asaro, "Liability Problem", *supra* note 3 at 191.

78. This is an important limitation on the foreseeability requirement: only the kind of resulting harm needs to be foreseeable, not the exact manner in which it actually occurs. However, as discussed above, see note 75, details of the harm's occurrence may be relevant where the description of tortious risk is specified narrowly. To circumvent this problem, one could construe the risk of deploying AAs broadly, as noted here in the text. I also discuss this strategy below in section IV.

79. To be clear, the problem here is not remoteness of damages—the *unforeseeable extent* of tortious causation of harm—which would implicate the thin-skull principle. Rather, AA-caused harm challenges the discovery of tortious causation of harm itself. The concern is that since AAs self-develop due to environmental inputs, resulting dangerous behaviour and harmful effects are not caused by tortious action at all. The thin-skull principle operates to make defendants liable for the full extent of damage where their tortious acts cause a kind of harm, the potential for which makes their acts wrongful, even if the extent of damages was not foreseeable. It still assumes that there is a tortious act—constituted by a foreseeable and unreasonable

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

to negligent use or design. Users should be liable for negligently using AAs in improper locations, performing insufficient maintenance, or carelessly manipulating their features—disregarding product warnings—causing malfunction and harm.[80] Likewise, manufacturers, designers, and sellers should be liable for distributing AAs that have demonstrably defective designs causing harm. However, at issue is a specific scenario: where AA-caused harm stems from AAs' emergence. In these instances, the harm results from misaligned algorithms due to inscrutable self-learning mechanisms, not simply traceable to legally defective design or negligent operation. It may be more-precisely termed *AA-caused harm*. A straightforward application of tort doctrine suggests that cost associated with such harm lies where it falls. Absent foreseeable and unreasonable risk of algorithmic misalignment and (the resulting type of) AA-caused harm, the basis in tort for imposing a duty to repair on any particular user or distributor appears to fall away. If this principled preclusion of liability strikes some lawyers as odd,

risk to the plaintiff (i.e. breach of duty of care)—causing harm, the potential for which made the act tortious (i.e. proximate causation). The argument in the text is that there are serious obstacles to viewing AA-caused harm as tortious causation of harm: first, if the deployment of AAs does not entail foreseeable and unreasonable risk of harm, it is not itself tortious, and second, even if a tortious act (in operating or programming the AA) is discovered, there may be further difficulty demonstrating that the tortious act caused the harm, rather than some other environmental input. Even a retreat to the (less stringent) directness standard of proximate causation, formulated in *In Re Polemis and Furness, Withy & Co*, [1921] 3 KB 560 (CA) [*Polemis*], does not simply resolve this problem. Although the *Polemis* approach finds liability for any harm that is a direct consequence of a tortious act—supplanting the need to demonstrate that the resulting harm is the kind of harm that that made the act tortious—it still assumes that a tortious act directly causes harm, which inevitably depends on the existence of tortious risk, conditioned by the notion of reasonable foreseeability. Consider this statement in *Polemis*, per Scrutton LJ: "To determine whether an act is negligent, *it is relevant to determine whether any reasonable person would foresee that the act would cause damage* .... But if the act would or might probably cause damage, the fact that the damage it in fact causes is not the exact kind of damage one would expect is immaterial, *so long as the damage is in fact directly traceable to the negligent act, and not due to the operation of independent causes having no connection with the negligent act*, except that they could not avoid its results" [emphasis added]. Accordingly, *Polemis* does not supplant the reasonable foreseeability determination; it is necessary to render conduct tortious. Moreover, to ground tort liability, the harm must be a direct consequence of the tortious act, and not due to independent causes that have no connection to the tortious act. A plausible ramification of the *Polemis* formulation is that where harm stems from unpredictable self-learning algorithms impacted by environmental inputs, such harm may be viewed as having an independent cause without connection to the tortious aspect of the act (even if an initial tortious act is identified), so will not ground liability.

80. For instance, in *Assiniboine*, *supra* note 74, the defendant invented an alternative, and exceedingly dangerous, means of starting an auto-toboggan which was not in the instruction manual. As a result, his son lost control of the machine and struck a gas-riser pipe in a nearby school. Because the pipe was fractured, gas escaped into the boiler room and was ignited by an open pilot light, causing explosion, fire and damage to the school building. Although the way the damage occurred—by fire and explosion—was not foreseeable, Dickson JA found that kind of physical damage that occurred was still foreseeable (due to impact of the auto-toboggan) and held the defendant liable. *Assiniboine* illustrates the limits of the foreseeability requirement: only the kind of resulting harm needs to be foreseeable, not the exact manner in which it actually occurs. (Note that the choice to adopt a broad description of risk is similar to *Hughes*, *supra* note 74.) At the same time, however, in *Assiniboine*, liability was traced to a specific tortious act, the dangerous manipulation and use of the auto-toboggan entailing foreseeable and substantial risk of harm.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

they will need to answer this question: on what theory may manufacturers, designers or users be held legally responsible?

This framing of the doctrinal issue reflects a particular theoretical view of tort law, as embodying a practice of corrective justice. My contention is that the doctrinal problem of AA-caused harm is due, in large part, to the fact that corrective justice is implicit in tort law, as its predominant conceptual and normative structure. It is, therefore, worthwhile to briefly highlight the foundational principles of corrective justice, as embodied in tort law, to demonstrate in what way AAs fit uneasily within tort's basic structure. As a practice of corrective justice, tort requires agent-specific reasons for liability: the plaintiff's right to recovery depends on a defendant's duty to repair. In this respect, tort liability responds to a specific scenario: where wrongful action materializes in loss to an object of the plaintiff's right. The locus of liability is wrongful loss; it is not just any loss or misfortune suffered by innocent victims, nor any unreasonable activity undertaken by tortious actors. Rather, tort liability operates upon the conjunction of a particular tortious action (i.e. breach of duty) with the causation of harm. This ordering is not inevitable, other systems of accident regulation are conceivable, indeed often endorsed. For instance, all innocent victims could recover from administrative compensation pools, funded by fines collected from negligent actors. Alternatively, mandatory insurance schemes can spread accident costs among a broad group of industry participants as part of an ongoing scheme that fairly allocates the burdens of enterprise in proportion to its benefits. Moreover, even under the prevailing litigation system where plaintiffs sue particular defendants to repair their losses, the norms determining liability could conceivably reflect concerns other than duty of care, wrongdoing and causation. Plaintiffs could make cases by proving that a particular defendant is best situated to payout the accident costs on bases of efficiency or comparative moral desert. However, under prevailing tort law, liability depends on proving this case: that a defendant harmed an object of the plaintiff's right in breach of a duty of care owed to the plaintiff. This includes three core features: first, a specific right-duty relation between a particular plaintiff and a particular defendant; second, wrongful conduct (breach of duty or misfeasance) by the defendant; and third, the defendant's wrongful conduct caused the plaintiff's loss (to an object of the plaintiff's right). These features may be explained as comprising a system of corrective justice: because the defendant's wrongful action caused loss to an object of the plaintiff's right, the defendant is liable to repair it. In this sense, tort is concerned with a version of outcome-responsibility: liability results from the attribution of a plaintiff's loss to a wrongful action, as its legally responsible cause. As a system of corrective justice, tort law implicates responsible human agency, grounding agent-specific duties to rectify harmful effects of wrongful action.[81] It does not merely shift accident costs amongst their participants pursuant to some allocative criteria, such as efficiency or merit.

81. Jules L Coleman, "The Practice of Corrective Justice" in David G Owen, ed, *Philosophical Foundations of Tort Law* (Clarendon Press, 1995) 53 at 66-69.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

The doctrinal integration of duty, wrongdoing and causation indicates that neither wrongful action nor causation, taken independently, provide grounds in tort to make a particular defendant liable to repair a plaintiff's loss. Tort liability responds to breaches of qualified duties of relational non-injury:[82] while there is no liability without causation of damage, such causation needs to occur in a particular way, as a materialization of wrongful action and with respect to a duty owed to the plaintiff. Since tort is constituted by the conjunction of duty, wrongdoing and causation, a successful account of tort cannot render any individual element—neither wrongdoing nor causation—foundational. The corrective justice account has explanatory power, as it effectively integrates duty, wrongdoing and causation. The idea of corrective justice is that individuals are constrained to act in accordance with a system of rightful interaction, as defined by legal actors' equal and reciprocal rights and duties of action.[83] In the world of corrective justice, liability follows only from defendants' actions that are inconsistent with this right-duty equilibrium, which necessarily implicates duty, wrongdoing and causation. Absent wrongdoing, causation of loss is not a tort: if the act is consistent with norms of rightful action, the causation of damage—though perhaps diminishing the plaintiff's welfare—is not inconsistent with, and does not infringe, the plaintiff's right.[84] Likewise, absent causation, a wrongful act is not a tort, as in this instance, there is no interference with an object of a plaintiff's right at all.[85] Finally, the doctrinal elements of duty and proximate cause specify the bilateral structure that integrates wrongdoing and causation. While the notion of duty forms the tort relation between a particular defendant and plaintiff as a legally significant unit, proximate cause confirms that the wrongdoing and causation are two stages in the existence of the same risky action, from its potential to cause a particular kind of harm to its realization as a completed tort, a wrongful interference with the plaintiff's right.[86] In this way, the defendant's action is deemed the legally relevant cause of the plaintiff's loss, grounding liability, the defendant's duty to repair the wrongful loss.

In construing the right-duty relation between legal subjects, tort liability—as a practice of corrective justice—reflects the formal equality in parties' standing as represented by their correlative (i.e. equal and reciprocal) rights and duties.[87] Tort law demarcates the scope of rightful action with respect to the parties' equal standing. This relates to the objective standard of reasonable care, which sets a level of reasonable risk imposition consistent with parties' equal rights to act in the world.[88] Any act of risk imposition that exceeds the reasonable level

82. Arthur Ripstein & Benjamin C Zipursky, "Corrective Justice in an Age of Mass Torts" in Postema, ed, *Philosophy and the Law of Torts* (Cambridge University Press, 2001) 214 at 218-19.
83. Ernest J Weinrib, "Causation and Wrongdoing" (1987) 63:3 Chicago-Kent L Rev 407 at 425-29.
84. Benson, *supra* note 62 at 766-70.
85. Ripstein, *supra* note 62 at 116.
86. Weinrib, "Causation and Wrongdoing", *supra* note 83 at 440-41.
87. *Ibid* at 426.
88. *Ibid* at 427-28.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

is wrongful. An act's wrongfulness does not depend on whether the defendant's subjective capacities of foresight and prudence empirically meet those of an 'ordinary person,' or whether the defendant could have indeed acted differently in the circumstances. Since the standard needs to reflect the equal standing of both interacting parties, it cannot depend on the defendant's subjective capacities. Rather, tort duties of reasonable care are reciprocal constraints on individuals' actions, as part of a system of cooperative social interaction. In this view, liability is a judgment of actions, not individuals. It evaluates whether individuals' actions are, or are not, consistent with the rightful actions of others; it does not judge individuals' inner-character, blameworthiness or actual lack-of-concern for others.[89] Through equal rights and duties—embodying an ideal of reciprocity—legal subjects need to honour fair terms of interaction and can be held legally responsible for their excessively risky (i.e. tortious) actions causing harm.

In sum, corrective justice is a theoretical framework that integrates a cluster of key concepts present in tort: bilateral duties of misfeasance, agent-specific grounds of liability, wrongdoing, causation, and the objective standard of care. As a practice of corrective justice, tort law constructs a normative framework for rightful interaction, and requires defendants to repair their wrongful losses. In this view, tort is not merely a species of regulatory law, a collection of incentives and liabilities aimed at fair and efficient allocation of accident costs between parties or across society more broadly. Certainly, tort law also achieves, at least in part, a range of social goals, such as deterring socially-harmful behaviour, reducing accidents or mitigating their costs, compensating faultless victims of unfortunate losses, and distributing the costs and benefits of industry in a more general sense. However, at its core, tort liability employs principles like duty, wrongdoing and proximate causation to attribute wrongful loss to responsible human agency. Tort's procedural and doctrinal principles suggest that its broader social achievements are only external by-products of its core practice of corrective justice, embodied in its rules. For example, in the world of tort, the social goal of compensating faultless victims is achieved only where a defendant is responsible in the legally relevant way. Likewise, defendants are not subject to deterrence measures or sanctions—such as fines—for engaging in unreasonably risky activities that, by luck, do not materialize in harm to a particular plaintiff to whom a duty is owed. Further, the law merely empowers plaintiffs to sue defendants for reparation if they want to do so; it does not automatically compel defendants to repair their wrongful harm, though it may advance social goals like deterrence or fair distribution of accident costs. Tort law is not simply aimed at *allocating* costs, the burdens of risky activity. Rather, tort law *attributes* plaintiffs' losses to particular defendants' actions on normative grounds, justifying agent-specific reparations by defendants.

Consequently, as a matter of tort law, the problem of AA-caused harm cannot be resolved solely by considering the fair allocation of AAs' accident costs. Tort

89. OW Holmes Jr, *The Common Law* (Little, Brown, 1881) at 108.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

analysis requires deeper treatment of the roles of human agency, responsibility, wrongdoing and causation in the context of AA-mediated relations. The selection of a particular defendant to repair a particular plaintiff's AA-caused-loss needs to reflect a normative link between the defendant's tortious action and the plaintiff's loss. In this respect, tort's corrective justice framework increases the complexity of discovering bases of tort liability for AA-caused. It must consider whether human or corporate manufacturers, designers, owners or users can be responsible for AA-caused harm in the legally relevant way, i.e. on the basis of duty, wrongdoing and causation. One may be tempted to bypass this doctrinal configuration and determine liability for AA-caused harm on the basis of distributive justice considerations—i.e. how to fairly and efficiently allocate accident costs between suppliers, users and victims of AAs—without diving into thorny notions of wrongdoing or outcome-responsibility. In this respect, positing various policy goals—such as incentivizing innovation, cost-justified reduction of accidents, compensating victims—and crafting liability standards to realize them may be a more fruitful undertaking. However, while this method is open to legislators to craft liability schemes for various kinds of AA-caused harm, tort analysis proceeds on different grounds. Tort liability for AA-caused harm necessarily implicates corrective justice, since it is implicit in tort's prevailing institutional and doctrinal framework—the bipolar structure of litigation, bilateral duties of misfeasance owed by particular defendants to particular plaintiffs, notions of wrongful action and responsible causation—embodying agent-specific bases of liability. Therefore, a successful doctrinal resolution to the problem of AA-caused harm needs to consider the relation between AAs' anticipated social roles and the presumptions of corrective justice that inform tort liability.

AA-caused harm is problematic for the theoretical structure of corrective justice. Since tort law demarcates and enforces norms of rightful interaction between particular defendants and plaintiffs, it presumes a juridical relation between two legal persons equally subject to tort's norms of interpersonal justice. As noted above, tort law assumes a bilateral relation as its subject: two legal persons situated symmetrically with correlative rights and duties, delineated by its doctrinal terms of fair interaction. The advent of AAs challenges this basic bilateral structure. Since AAs are not legal subjects bearing tort duties, AA-caused loss—where not simply traceable to programming, instruction or manipulation by human programmers or users—does not stem from pertinent legal agency, much like loss due to natural disasters. Although AAs are purposely deployed by humans—unlike natural disasters—victims may be left without an avenue for compensation in tort. A formalistic application of tort law suggests that all such losses lie where they fall, like damage due to natural disasters. However, this position is intuitively unsatisfactory. First, it is at odds with intuitions about fair allocation of burdens and benefits of sophisticated technologies—notions of distributive justice—that are not accounted for in pure tort analyses. Further, it is arguable that this result does not properly reflect tort law's norms of interpersonal justice. It would be odd if utilization of AAs could preclude tort liability, whether in principle or practice. The intuition is that standard application of tort law, without

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

modification, fails to adequately capture the nature of AA-mediated human interaction. The problem with the standard application is that it implicitly views AAs as ordinary tools. However, AAs' functional independence and distinctive utility suggests that they are better perceived as quasi social actors, with capacity to independently provide benefits and cause harm. Letting all AA-caused losses lie where they fall reveals a fundamental incongruity: while suppliers and users deploy AAs for their benefit, enabling AAs to independently cause harm, they are immune to tort liability because the AAs—which are not legal subjects—are the relevant social actors causing the harm. Whereas personally performing tasks causing harm could potentially produce liability, individuals can avoid tort liability altogether by deploying AAs to perform the tasks instead.

The ostensible incompatibility between tort law, corrective justice and AA-caused harm needs to be properly demonstrated, requiring more-probing analysis of tort doctrine. To this end, the following section examines several doctrinal strategies to ground tort liability for AA-caused harm under the laws of negligence and strict liability. It concludes, however, that these strategies are largely ineffective. Absent modification or reconception of tort principles, AA-caused losses likely lie where they fall.

IV. Doctrinal Strategies to Discover Tort Liability for AA-Caused Harm

1. The Law of Negligence and AA-Caused Harm

The scenario at issue is AA-caused harm: since AAs are designed to self-develop their algorithms through processes of trial-and-error, their final algorithms may be misaligned with programmers' and users' actual goals, with potential to cause accidents. A determination of tort liability for AA-caused harm, under the rubric of negligence, requires a doctrinal basis to impute AA-caused harm to human (or corporate) wrongdoing. However, these accidents are not easily traceable to human error, as misalignment can occur through AAs' internal learning processes unknowable to programmers or users. Where harm results from AAs' misaligned algorithms due to inscrutable self-learning mechanisms, it may be exceedingly difficult, perhaps impossible, to trace connections between various environmental inputs and ensuing changes in AAs' algorithms and behaviours. There are then two major hurdles to discovering a basis of liability under the law of negligence: The first is the foreseeability of real and substantial risk, a formal constraint that informs the boundaries of duty, tortious conduct and proximate causation. As discussed above, defendants owe duties, and are obligated to modify conduct, only with respect to foreseeable risks to particular plaintiffs. Likewise, tortious causation occurs only if plaintiffs' harm is both attributable to, and within the scope of, tortious conduct involving foreseeable and unreasonable risk. If AAs' algorithmic developments and harmful outward effects are unforeseeable, these conditions of liability may not be met. The second is the structure of the tort relationship, constituted by two legal subjects, whereby

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

one actor wrongfully causes loss to the other. If AA-caused harm is more precisely attributable to the AA itself—which is not a legal subject—it may be doubted that such harm is a matter of tort law.[90]

As a general rule, the fact of harmful accident is not evidence of negligence; the plaintiff needs to prove negligence on a balance of probabilities, it is not presumed.[91] However, circumstantial evidence may be sufficient to justify an inference of negligence.[92] One kind of circumstantial evidence, termed *res ipsa loquitur*—"the thing speaks for itself"—permits inferences of negligence in circumstances where, in the ordinary course of things, the accident would not have occurred if the defendant had taken proper care.[93] However, in the context of AA-caused harm, the fact of accident or dangerous activity does not permit a strong inference of negligence, perhaps none at all. Since emergent AAs self-develop through interaction with their environments—entailing unforeseeable input—unpredictable output, including error, is expected to occur in isolated instances. It cannot be said that in the ordinary course of AAs' development and deployment, accidents would not occur unless their users or programmers failed to take proper care. Due to AAs' emergent features, isolated accidents are expected even for well-trained and reliable systems. Accordingly, in the context of AAs, the fact of accident is neutral evidence without further proof of negligence.

One possible strategy for plaintiffs is to analogize instances of AA-caused harm to cases of liability for actions of third parties. Claimants may argue that defendants should be liable for contributing to conditions or initiating a process that leads to harm, though directly caused by another entity, the AA. There are notable instances of liability despite the presence of an intervening cause. In *Weirum v RKO General*, a defendant radio station broadcasted a prize offer to the first person who locates a particular disc jockey travelling in a conspicuous red car; since the broadcast (foreseeably) induced two drivers to dangerously chase the disc jockey and cause a fatal accident, the radio station was liable.[94] Likewise, commercial establishments serving alcohol may be liable to third-party road-users for harmful accidents caused by their intoxicated customers.[95] In these cases, defendants are apparently liable for facilitating a dangerous condition, whereby other actors directly cause harm. Analogously, the argument goes, human deployers of AAs can be held liable for facilitating conditions leading to AA-caused harm.

However, this argument likely fails. Liability does not follow simply from initiating a process that leads to harm in a purely causal sense. Liability attaches

90. This analysis assumes AAs are not legal persons, i.e., rights-and-duties-bearing subjects with legal capacity to comprise tort relations. For the purposes of this essay—situating AA-caused harm with negligence and strict liability frameworks—my analysis assumes AAs are not legal persons. This is simply AAs' default legal position absent explicit recognition of their personhood in a legal system.

91. William L Prosser, *The Law of Torts*, 4th ed (West, 1971) at 211.

92. *Ibid*.

93. *Fontaine v British Columbia (Official Administrator)*, [1998] 1 SCR 424 at paras 17, 23-27.

94. *Weirum v RKO General, Inc*, 15 Cal (3d) 40 (1975).

95. *Stewart v Pettie*, [1995] 1 SCR 131 at para 28.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

to a defendant's unreasonably dangerous (i.e. tortious) conduct; it does not attach to conduct that merely provides an opportunity for other actors to endanger the plaintiff.[96] For this reason, in contrast to commercial establishments, private social hosts serving alcohol do not usually owe duties to protect members of the public from negligent actions of intoxicated guests.[97] The legal distinction between commercial and private hosts arguably relates to whether the hosts' conduct is implicated in the creation or exacerbation of the risk. Commercial hosts, by definition, encourage alcohol consumption as part of doing business. By contrast, as a general matter, social hosts merely provide a place for people to meet and possibly consume alcohol if they so choose. Absent active implication in the creation of risk—such as encouraging visibly intoxicated guests to further imbibe—social hosts are not usually responsible for their guests' independent alcohol-consumption choices.[98] The distinction between commercial and social hosts is certainly not self-evident; classifying the cases differently—where both commercial and social hosts are subject to the same standard, whether under a duty or otherwise—is conceivable.[99] Nonetheless, these cases establish how to think about liability in cases of intervening-cause: the defendant's action needs to create a dangerous condition that leads to the plaintiff's harm, not merely create a condition that provides others with an opportunity to impose danger on the plaintiff.[100] In this way, the form of inquiry in intervening-cause cases follows a proximate cause analysis: liability responds to an initial tortious act that is likely to lead, and in fact leads, to the ensuing action of the third-party that directly causes harm.[101] The inquiry focuses on drawing a link between the defendant's initial tortious risk and the resulting harm.[102] The requisite link can be formed or

96. Ripstein, *supra* note 62 at 97.
97. *Childs v Desormeaux*, [2006] 1 SCR 643 at paras 44-45.
98. *Ibid*.
99. Ripstein, *supra* note 62 at 98.
100. *Ibid* at 97.
101. *Home Office v Dorset Yacht Co Ltd*, [1970] AC 1004 (HL) [*Dorset Yacht*].
102. This form of inquiry is also observable in a recent Supreme Court of Canada case, *Rankin (Rankin's Garage & Sales) v JJ*, [2018] 1 SCR 587 [*Rankin*]. In *Rankin*, two teenagers, both minors, stole an unlocked car form an unsecured commercial car garage property. Though he had never driven before (and did not have a driver's license), one of the minors drove onto a highway and crashed the car, causing the other to suffer a catastrophic brain injury. At issue was whether the commercial garage owner owed a duty of care to the injured teenager when securing cars stored in the garage. The Supreme Court stated that the proper question is whether the kind of harm suffered—personal injury—was reasonably foreseeable to the car garage owner. The Court emphasized that physical injury would be reasonably foreseeable only if there was risk of theft *and* that, upon theft, there was risk the stolen car would be driven dangerously. While the Court noted that risk of theft by minors entails risk of reckless driving and would, in theory, link the garage owner's failure to secure the cars to the kind of harm suffered, it found that there was insufficient evidence of specific risk of theft by minors: minors did not frequent the garage area at night, nor were they especially engaged in stealing cars. Accordingly, the Court held that the resulting injury was not a reasonably foreseeable outcome of the failure to secure the cars, negating the tort duty in the circumstances. This case offers an illustration of what is at stake in determining liability where there is an intervening cause: the resulting harm must be the kind of harm that made the initial act tortious. To this end, the reasonable foreseeability of the intervening act—in this case, theft involving risk of dangerous driving—is crucial. Since failing to secure the car did not constitute tortious risk (of theft involving risk of dangerous driving), the resulting harm was not linked to an initial tortious act.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

broken by both human or other natural intervening-causes; it depends entirely on the likelihood of the intervening-cause occurring as a result of the initial negligent act.[103] Consequently, in the context of AA-caused harm, liability depends on whether the AA's action forms the requisite link between an initial tortious act and resulting harm.[104]

Consider an illustrative case, *Home Office v Dorset Yacht*, where a defendant borstal detention centre was held liable for damage to a plaintiff's boat caused by borstal boys in its custody.[105] This case offers an interesting analogy to AA-caused harm: just as the defendant detention centre has special responsibility for actions of borstal boys in its custody—even though they are independent human actors, beyond the defendant's full control—likewise, human (or corporate) defendants may have special responsibility for using or distributing AAs, even though they do not have complete control over AAs' effects (or doings), entailing a degree of functional independence. In *Dorset Yacht*, the borstal boys seized and damaged the plaintiff's boat in the course of escaping from the detention centre, under the negligent supervision of borstal officers. In finding liability, Lord Diplock held that the officers' negligent supervision breached a duty to the plaintiff because the plaintiff's boat was nearby, and there was real and foreseeable risk that the borstal boys would seize and damage it in the course of escaping. Liability was rooted in the initial tortious conduct of the borstal officers, their negligent supervision. Since the borstal boys' conduct (the intervening-cause) and the resulting damage to the boat fell within the scope of the officer's negligence, the intervening cause did not break the chain of causation linking the initial negligent supervision and the resulting property damage.[106] Accordingly, the

103. It is not entirely clear whether there is a legal distinction between human intervening causes and other natural intervening causes. For example, the dissent in *Bradford v Kanellos*, [1974] SCR 409, 40 DLR (3d) 578 [*Bradford*] at 582 adopts a formulation stating there is no categorical distinction between forces of nature and human action, even consciously controlled. However, as noted below in the text, leading UK cases indicate that forming proximate causation is subject to a higher standard in cases of human intervening causes, in which the chain of causation is broken unless their intervening actions are likely, even inevitable. See *Dorset Yacht*, supra note 101 per Lord Reid; also see *Lamb v London Borough of Camden*, [1981] QB 625 (CA), per Oliver LJ. Nevertheless, while the standards for forming proximate cause may differ, the form of argumentation is the same for natural and human intervening causes: the intervening cause needs to be construable as forming a link between an initial wrongdoing and resulting harm.

104. Related to the previous note, an interesting question arises as to whether AAs' actions need to be likely (or inevitable)—not just foreseeable or "not abnormal"—to form proximate causation between initial wrongdoing and the resulting AA-caused harm, akin to human intervening causes.

105. *Dorset Yacht*, *supra* note 101.

106. The idea is that liability is rooted in the defendant's own tortious act. It is not properly construed as responsibility for the intervening acts of another. This is nicely illustrated in Lord Reid's framing of the issue: "Even so it is said that the respondents must fail because there is a general principle that no person can be responsible for the acts of another who is not his servant or acting on his behalf. *But here the ground for liability is not responsibility for the acts of the escaping trainees; it is liability for damage caused by the carelessness of these officers in the knowledge that their carelessness would probably result in the trainees causing damage of this kind.* So the question is really one of remoteness of damage. And I must consider to what extent the law regards the acts of another person as breaking the chain of

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

liability inquiry involves finding tortious conduct and proximate causation; the mere initiation of conditions leading to harm in a causal sense is insufficient to ground liability.

Moreover, the analogy between AA-caused harm and human intervening-cause cases arguably works against the interests of victims of AA-caused harm. According to an established line of case law, to form proximate causation between initial tortious conduct and resulting harm, a human intervening-act needs to be a likely, even inevitable, outcome of the initial tortious act, not just foreseeable.[107] Therefore, it is incorrect to frame instances of liability for harm caused (directly) by third parties as legal responsibility for acts of others. Since liability occurs only if the intervening acts are natural and likely outcomes of initial tortious conduct, defendants are liable for their own wrongful actions which are deemed the responsible causes of harm due to their chain of probable causal effects. Intervening-cause cases neither supplant the tortious conduct inquiry, nor do they deem as tortious the initiation of processes that are merely causally traceable from resulting harm. If AAs' actions are indeed akin to human intervening causes, AAs' dangerous actions need to be likely or inevitable to link their resulting harm to an initial tortious human or corporate action, a standard not easily met. Cases of liability for actions of third parties are then unlikely to support liability for AA-caused harm, absent further discovery of tortious conduct and proximate causation.

Accordingly, victims of AA-caused harm need to demonstrate that human tortious action created a dangerous risk leading to AA-caused harm. However, even if victims successfully establish the existence of negligence, they may still face

causation between the defendant's carelessness and the damage to the plaintiff" [emphasis added]. *Ibid* at 1027.

107. *Ibid* at 1030: "[W]here human action forms one of the links between the original wrongdoing of the defendant and the loss suffered by the plaintiff, the action must at least have been something very likely to happen if it is not to be regarded as a *novus actus interveniens* breaking the chain of causation. I do not think that a mere foreseeable possibility is or should be sufficient, for then the intervening human action can more properly be regarded as a new cause than a consequence of the original wrongdoing." See also *Lamb v London Borough of Camden,* [1981] QB 625 (CA) at 644, per Oliver LJ, commenting on Lord Reid's formulation: "I would respectfully regard Lord Reid's test as a workable and sensible one, subject only to this; that I think that he may perhaps have understated the degree of likelihood required before the law can or should attribute the free act of a responsible third persons to the tortfeasor . . . . It may be that some more stringent standard is required. There may, for instance, be circumstances in which the court would require a degree of likelihood amounting almost to inevitability before it fixes a defendant with responsibility for the act of a third party over whom he has and can have no control." Arguably, this approach is also followed by the Supreme Court of Canada in *Bradford*, *supra* note 103. In *Bradford*, the defendant negligently failed to clean a restaurant grill leading to fire, which was subsequently extinguished. However, the fire-extinguisher made a hissing noise, prompting a costumer to yell that gas was escaping and that there was danger of explosion, causing customers to panic, rush to the doors and injure the plaintiff in the process. The Court held that the plaintiff's harm caused by third-party customers was not within the scope of the risk constituted by the negligent failure to clean the grill. The dissent argued, however, that the customer's panic-inducing exclamation of explosion was an "utterly foreseeable" response to the negligent fire, part of "the natural consequence of events leading inevitably to the plaintiff's injury." *Ibid* at 582. In this respect, the dissent seems to ground liability in the *inevitability* of the intervening act.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

difficulty proving that the AA's dangerous behaviours stemmed specifically from the negligent act, rather than some unknown environmental input, which possibly breaks the chain of causation. To increase the chances of a successful negligence case, plaintiffs could propose a broad description of tortious risk entailed in distribution or deployment of AAs. They may argue that the distribution or deployment of AAs is itself a tortious risk, as it entails risk of accidents, at least in isolated cases, due to potential misalignment. In their view, suppliers or users should be responsible for creating allegedly dangerous conditions whereby AAs interact independently with other persons and property with potential to cause harm.

However, this approach is problematic: it reflects a strict liability standard, not negligence. Since the judgement of tortious conduct is attributed to risks inherent in deploying AAs—rather than risks associated with particular ways of deploying AAs—the deployment of AAs is itself deemed tortious. While this standard does not necessarily lead to absolute liability—there is liability only in cases where the resulting harm is of a kind that makes deployment tortious—it does impose liability without providing defendants opportunity to modify their conduct. Crucially, applying a strict liability standard to AA-caused harm requires a separate analysis. It must be demonstrated that deployment of AAs is the kind of activity that should be subject to strict liability, analogous to recognized strict liability categories, such as abnormally hazardous activities or keeping wild animals. This is the subject of section IV.2 below.

In sum, the law of negligence does not provide a clear basis of liability for AA-caused harm. AAs' emergence and functional independence precludes designers or users from having full control over AAs to prevent AA-caused harm. AA-caused harm is not easily traceable to particular human wrongdoing, but more directly attributable to internal processes of AAs themselves. As an evidentiary strategy, *res ipsa loquitur*, is unlikely to assist plaintiffs: the fact of accident is not indicative of negligence, as misalignment and harmful accidents can occur due to AAs' emergence. Case law establishing liability for harmful actions of third parties is also unlikely to assist plaintiffs: since harmful effects of AAs is not an inevitable outcome of their deployment in any particular instance, it is not linked to initial tortious action. And to describe the tortious risk as one inherent to the deployment of AAs, effectively moves toward a strict liability model, beyond negligence. Accordingly, under the law of negligence, the costs of AA-caused harm appear to lie where they fall.

This outcome is not surprising: tort law embodies corrective justice, assuming a relation between two legal subjects, whereby one subject wrongfully causes loss to the other, a breach of duty that infringes a right. As a version of outcome-responsibility, tort law identifies a particular responsible legal agent to repair the plaintiff's loss. AA-caused harm does not fit neatly within this framework: First, harm that is attributable to AAs cannot ground liability, as AAs are not legal subjects capable of breaching tort duties. Further, attempts to attribute AA-caused harm to human wrongdoing fail because AAs act with functional independence, breaking the chain of proximate causation traceable to legal agency.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

2. Strict Liability and AA-Caused Harm

I now turn to assess whether the distribution or use of AAs should be subject to a strict liability standard. The pertinent issue is whether distributing or deploying AAs is itself tortious, i.e. whether it necessarily entails real and substantial risk to some plaintiff with respect to some kind of harm. This requires judgment about how societal participants should view AAs: Are AAs dangerous items with tendencies to escape from human control and cause mischief in the event of escape? Is utilizing AAs more risky than ordinary activities, comparable to keeping wild animals, blasting or transporting gasoline?

To begin, each of the recognized categories of strict liability—holding artificial accumulations of liquid,[108] blasting,[109] transporting gasoline,[110] keeping wild animals[111]—follow the same basic structure: a defendant carries on an activity that entails real and substantial danger to particular plaintiffs if the potentially-dangerous object of the activity escapes the defendant's control. Moreover, each of these activities involve an object that has a tendency to escape: percolating water, explosions and wild animals have tendencies to escape human control, and expectedly cause mischief if they do.[112] The point is that, in the law's view, these objects cannot be controlled completely, so cannot be made safe, rendering defendants' precautions irrelevant.[113] Since these activities necessarily entail risk imposition beyond those assumed in ordinary patterns of interaction, plaintiffs do not need to show that defendants failed to exercise proper care.[114] Even the most conscientious user of explosives, transporter of gasoline, or keeper of wild animals is doing something too risky, i.e. imposing risk exceeding the level of risk present in ordinary activities.[115] Since these activities are necessarily too dangerous in a specific way to a specific plaintiff, they constitute tortious conduct.[116] At issue is whether utilizing AAs is sufficiently analogous to these strict liability categories: is it is an activity carrying too much inherent risk to enable the preclusion of tort liability through exercising reasonable care?

Victims of AA-caused harm should aim at this argument: The mere deployment of AAs entails inherent risk of misalignment and possibility of error, potentially resulting in harm to persons or property. Since AAs have emergent capacities to learn from environmental inputs and adjust their algorithms in ways unknowable to designers or users, AAs' outward effects are characteristically unpredictable, undermining the possibility of full human control. In this sense, the argument goes, AA-caused harm is analogous to paradigmatic instances of strict liability. The risks of misalignment, error and harm due to emergence

108. See, e.g., *Rylands v Fletcher* (1868), LR 3 HL 330 (HL).
109. See, e.g., *Spano v Perini Corp*, 25 NY (2d) 11 (NY 1969).
110. See, e.g., *Siegler v Kuhlman*, 502 P (2d) 1181 (Wash Sup Ct 1972) at 454 [*Siegler*].
111. See, e.g., *Behrens v Bertram Mills Circus* [1957] 2 QB 1.
112. Ripstein, *supra* note 62 at 141.
113. *Ibid* at 142.
114. *Ibid*.
115. *Ibid* at 143.
116. *Ibid* at 142.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

is comparable to the tendency of explosives or wild animals to escape human control and cause mischief. Since the deployment of AAs involves inherent risk of uncontrollable misalignment—akin to the tendency of explosives or wild animals to escape—AAs must be deployed at deployers' peril, inviting strict liability for resulting harms due to misaligned (i.e. escaping) algorithms.

Moreover, where harm results from misaligned algorithms due to inscrutable self-learning mechanisms, there may be little evidence, if any, of the accident's precise cause, i.e. whether it was caused by particular human or environmental data input.[117] Plaintiffs will then have undue difficulty proving that misaligned algorithms resulted from human negligence. Therefore, the argument continues, AA-caused harm is analogous to transporting gasoline and blasting activities which risk destroying relevant evidence of the causes of accident, thereby supporting the imposition of strict liability.[118]

A strict liability standard for deployment of AAs does not necessarily lead to absolute liability in all instances of AA-caused harm. As a requirement of proximate causation, the AA-caused harm needs to be of a kind that would foreseeably result from misalignment for the particular kind of activity the AA undertakes. At this early point, lacking experience with AA activity, it is difficult to consider which kinds of harm are especially risky for particular AAs. However, experience with AA activity may enable assessment about which kinds of harm are real risks—and which are not—for a particular AA. For example, in the context of automated vehicles, users may assume strict liability for road accidents falling within the scope of driving errors, such as a failure to identify a snow-covered stop sign. However, strict liability may not apply to accidents that fall outside the usual scope of driving. For instance, if a car is programmed to experiment with energy efficiency and starts its engine in a garage to recharge its battery causing a passenger's death by carbon monoxide, the resulting harm may fall outside the scope of driving-related harms.[119] The point is that strict liability for AA-caused harm takes this form: deployers are strictly liable where AAs cause harm falling within the scope of their inherent risk, thereby conforming to the structure of proximate cause.

In sum, the deployment of AAs inherently risks misalignment, error and harm. It also places plaintiffs in precarious positions, lacking evidence needed to prove negligence. Finally, strict liability for AA-caused harm is limited to harms that fall within the scope of an AA's risk (it is not absolute liability). The deployment of AAs is then ostensibly a legitimate candidate for strict liability.

---

117. See Part II above.

118. *Siegler*, *supra* note 110 at 454-55: "Dangerous in itself, gasoline develops even greater potential for harm when carried as freight .... It is quite probable that the most important ingredients of proof will be lost in a gasoline explosion and fire .... As a consequence of its escape from impoundment and subsequent explosion and ignition, the evidence in a very high percentage of instances will be destroyed, and the reasons for and causes contributing to its escape will quite likely be lost in the searing flames and explosions."

119. This example is in Calo, "Robots as Legal Metaphors", *supra* note 4 at 230.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

There is also a worthy counter-argument: strict liability for AA-caused harm misapprehends the level of danger posed by AAs. It conveys a problematic view of AAs: that they are dangerous, akin to wild animals and explosives, and best kept controlled under leash. Strict liability for AA-caused harm implies that AAs pose a degree of risk to plaintiffs exceeding levels of risk typically assumed in ordinary patterns of interaction. However, the counter-argument goes, it is mistaken to construe AAs as exceedingly dangerous. AAs may be less risky than the human actors they replace. For example, automated vehicles are likely to be significantly safer than conventional vehicles due to the absence of human-driver-error.[120] For this reason, Ryan Abbott recommends that a negligence standard apply to harms caused by autonomous machines—instead of strict liability—to incentivize the adoption of automations that reduce accidents, resulting in net safety gains.[121] If deploying AAs is expected to reduce risks associated with human activities, and thereby enhance consumer safety, it should not be categorized as an abnormally dangerous activity. It is dissimilar to keeping wild animals or using explosives, which involve characteristic risk exceeding the level of risk assumed in ordinary patterns of interaction.

Moreover, the counter-argument continues, the analogy to wild animals and explosives—implying AAs have an equivalent dangerous propensity to escape human control and cause mischief—is inaccurate. AAs are deliberately designed and deployed to interact with other persons and property. They are not supposed to be kept under strict control, unlike explosives or wild animals. AAs' functional autonomy is a feature, not a flaw; it underpins their capacities to perform complex tasks for human deployers. In this sense, the analogy to keeping wild animals or using explosives—entailing dangerous tendencies to escape and do mischief—is not compelling. In the context of keeping wild animals and using explosives, the tendency to escape is a source of undue risk; upon escape, they are prone to cause damage. Since these activities entail risk of *dangerous escape,* they are essentially tortious. Users of explosives and owners of wild animals are therefore under strict duties to keep their explosives and animals under control. By contrast, AAs' emergence is actually their source of value, intentionally utilized by humans. Since AAs' emergence is not exceedingly risky, deploying AAs should not be deemed tortious, i.e. it should not be subject to strict liability.

We have then two conflicting theories about whether deploying AAs should be subject to strict liability in tort. Notably, these arguments highlight an ambiguity in the analogy to the classes of activities subject to strict liability. The ambiguity relates to the nature of AAs' escape: Is their potential misalignment the pertinent escape? Alternatively, is their emergence the pertinent escape? In the first theory, arguing for strict liability, the pertinent escape is potential misalignment. In this view, deploying emergent AAs carries risk of misalignment, an escape that is dangerous, akin to dangerous propensities associated with

120. Bryant Walker Smith, "Automated Driving and Product Liability" (2017) Michigan State L Rev 1 at 15.
121. Abbott, *supra* note 18 at 22.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

explosives and wild animals. However, in the second theory, arguing against strict liability, the pertinent escape is AAs' emergence. In this view, the escape associated with AAs is not exceedingly dangerous, in contrast to the dangerous tendencies-to-escape associated with explosives and wild animals.

Arguably, the second theory is more compelling. To review, common law strict liability applies to activities involving objects that: (a) have escaping tendencies; *and* (b) are dangerous upon escape. Each of the paradigmatic classes of strict liability—wild animals, percolating water, tanks of gasoline and explosives—involve objects that have tendencies to escape human control *and* are dangerous if they do, in fact, escape. An apt analogy to these classes of ultra-hazardous activities would need to show that AAs have tendencies to escape human control *and* that they are exceedingly dangerous when they escape. My contention is that AAs should not be characterised this way. If the pertinent escape is AAs' emergence, then it is true that AAs have a tendency to escape. However, this kind of escape (i.e. emergence) is not exceedingly risky; in fact, it may result in net safety gains. If the pertinent escape is potential misalignment, then it is true that the escape, if it does occur, is dangerous. However, the risk of this kind of escape (i.e. misalignment) is not likely to occur in any particular instance, so it cannot be said that AAs have a tendency to escape. Accordingly, no matter how the pertinent escape is construed, AAs are arguably not analogous to the classes of activities subject to strict liability. If AAs are not exceedingly risky in relation to ordinary background risks, their deployment should not be deemed tortious, and it should not be subject to strict liability.

This is not to deny that isolated incidents of misalignment, error and harmful accidents are expected to occur. There may also be good policy reasons to compensate victims of unavoidable accidents associated with AAs. However, this would seemingly need to be implemented in the form of statutory compensation, not through tort law, at least as traditionally conceived.[122] The fact that there will be occurrences of unavoidable AA-caused harm does not render the deployment of AAs abnormally dangerous. In the absence of undue risk, there is no tortious conduct; in the absence of tortious conduct, there is no agent-specific reason to impose strict liability on a particular defendant.

3. A Note on US Strict Products Liability

I now turn to briefly address the role of strict products liability for AA-caused harm. Strict products liability provides limited assistance to determining tort liability for AA-caused harm, pursuant to corrective justice. First, strict products liability is a uniquely American regime.[123] In Canada, as a general matter, the law of negligence applies to products liability (that is, causes of action against

122. For example, in Quebec, there is statutory compensation for injuries due to vaccinations "regardless of responsibility". *Public Health Act*, CQLR, c S-2.2 at s 73.

123. John CP Goldberg & Benjamin Zipursky, "The Easy Case for Products Liability Law: A Response to Professors Polinsky and Shavell" (2010) 123:8 Harv L Rev 1919 at 1919.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

manufacturers, designers and sellers).[124] Moreover, the rationales for strict products liability are "more instrumental than doctrinal."[125] It is primarily a tool to promote enterprise liability, loss-spreading, efficient deterrence and compensation.[126] In this respect, strict products liability does not embody corrective justice. Finally, even in the United States, products liability for design defects—such as software and algorithm defects—is governed by principles of fault, not strict liability.

Strict products liability originated in the mid-20th century in the United States. From the mid-1960's to the mid-1980's, strict liability for product related accidents was widespread.[127] In a seminal case, *Escola*, Justice Roger Traynor held that a manufacturer should incur "an absolute liability when an article that he has placed on the market, knowing it is to be used without inspection, proves to have a defect that causes injury to human beings."[128] According to Justice Traynor, manufacturers can test products in ways consumers cannot.[129] Moreover, manufacturers can acquire insurance for product-related-harm as a cost of doing business, and widely distribute accident costs among the public by adjusting product prices.[130] For these reasons, it was held more preferable for accident costs to be internalized by manufacturers than by consumers. Nearly twenty years later, in *Greenman*, Justice Traynor, writing for a unanimous court, re-affirmed strict products liability, stating: "The purpose of such liability is to insure that the costs of injuries resulting from defective products are borne by the manufacturers that put such products on the market rather than by the injured persons who are powerless to protect themselves."[131] Accordingly, in its original manifestation, strict products liability reflected concerns of enterprise liability, an imperative to shift accident costs to manufacturers, away from injured consumers. The subject of strict products liability was accident costs and redistribution, not wrongful conduct and corrective justice.

Justice Traynor's approach to products liability was enshrined in the *Restatement (Second) of Torts*, which provides that commercial sellers are strictly liable for harm caused by their defective products, even if they exercise "all possible care in the preparation and sale" of their products.[132] Nevertheless, the doctrine of strict products liability had its limits. Courts agreed that manufacturers and sellers should not be insurers of a product's safety, and that products do not need to be "perfectly safe," "accident-proof," nor the "ultimate in safety."[133] Rather, manufacturers and sellers were liable only for *defective* products, i.e.

124. See *Phillips v Ford Motor Co of Canada Ltd*, [1971] 2 OR 637 (CA), per Schroeder JA.
125. Goldberg & Zipursky, *supra* note 123 at 1919.
126. *Escola v Coca Cola Bottling Co of Fresno*, 24 Cal (2d) 453 (Cal 1944), per Traynor J at 461-62 [*Escola*].
127. David G Owen, "The Evolution of Products Liability Law" (2007) 26:4 Rev Litig 955 at 977.
128. *Escola*, *supra* note 126 at 461.
129. *Ibid* at 461-62.
130. *Ibid* at 462.
131. *Greenman v Yuba Power Products, Inc*, 377 P (2d) 897 (Cal 1963) at 901 [*Greenman*].
132. *Restatement (Second) of Torts* § 402A(2)(a) (1965).
133. Owen, *supra* note 127 at 976-77.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

where something was wrong with the product.[134] To this end, courts devised tests to determine product defectiveness, including consumer-expectation and risk-utility (cost-benefit) tests.[135] By the late-1980's, strict products liability was declining, as courts allowed defences based on the state of applicable "scientific and technological art at the time of a product's manufacture and sale."[136] Crucially, by the late-1990's, the time of writing the *Restatement (Third) of Torts*, strict products liability was mostly overturned.[137] While products liability is often still termed 'strict,' this is largely due to it doctrinal formulation, in which liability turns on a product's defectiveness (i.e. the product's dangerous condition), rather than the wrongful conduct of its manufacturers, designers or sellers.[138] However, in a functional sense, since products liability applies only to *defective* products, it mostly turns on notions of fault.[139] In particular, to determine whether a product is defective, courts use principles of reasonableness, balancing factors like product usefulness, safety, cost and practicality.[140] The shift from strict liability to fault is reflected in the *Restatement (Third) of Torts*.

The *Restatement (Third) of Torts* provides that distributors of defective products are liable for harm proximately caused by product-defects.[141] In particular, the *Restatement* sets out three categories of product defects.[142] The first category relates to *manufacturing defects*: where a product's construction departs from its intended design.[143] Distributers are strictly liable for manufacturing defects that proximately cause harm, even if "all possible care was exercised in the preparation and marketing of the product."[144] However, this is the only products liability category that is currently subject to a strict liability standard. The second category relates to *design defects*: where a product's foreseeable risk of harm "could have been reduced or avoided by the adoption of a reasonable alternative design," and the "omission of the alternative design renders the product not reasonably safe."[145] The third category relates to *warning defects*: where a product's foreseeable risk of harm "could have been reduced or avoided by the provision of reasonable instructions or warnings," and the "omission of the instructions or warnings renders the product not reasonably safe."[146] Notably, the *Restatement* also adapts the evidentiary principle, *res ipsa loquitur*, to the context of products liability. It may be inferred that a plaintiff's harm was caused by a product defect where the harm is "of a kind that ordinarily occurs as a result of product defect,

134. *Ibid* at 977.
135. *Ibid*.
136. *Ibid* at 979.
137. *Ibid* at 979-82.
138. Goldberg & Zipursky, *supra* note 123 at 1923-24.
139. Owen, *supra* note 127 at 980-81.
140. *Ibid*.
141. *Restatement (Third) of Torts: Products Liability* § 1 (1998).
142. *Ibid* at § 2.
143. *Ibid* at § 2(a).
144. *Ibid*.
145. *Ibid* at § 2(b).
146. *Ibid* at § 2(c).

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

and was not, in the particular case, solely the result of causes other than product defect existing at the time of sale or distribution."[147]

These three liability standards "provide the foundation for most modern products liability law."[148] While manufacturing defects are subject to strict liability, design defects and warning defects are defined in terms of negligence.[149] In particular, liability for design defects requires the following conditions: (a) The product is defective, as the failure to adopt a reasonably alternative design renders the product unsafe; (b) the ordinary use of the defective product entails foreseeable risk of harm; and (c) the product defect is the proximate cause of the plaintiff's harm. Liability for warning defects follows the same structure.

Let us return to the problem of AA-caused harm. Strict products liability applies to AA-caused harm only in instances of manufacturing defects. Manufacturers, designers and sellers are exposed to strict liability if: (a) the construction of an AA departs from the intended design, constituting a manufacturing defect; and (b) the defect causes harm. This doctrinal possibility does not resolve our issue, however, where AA-caused harm is due to software design—including self-learning algorithms—and not linked to specific construction error. It may be possible to construe software and algorithmic errors as potential design defects. However, as noted, this turns on a negligence standard, constrained by notions of foreseeability and proximate causation.[150]

This is not to deny the relevance of products liability for AA-caused harm in a more general sense. Since AAs are artifacts—designed, manufactured and sold—their harmful effects implicate their designers, manufacturers and sellers. For this reason, in the automated driving context, Bryant Walker Smith predicts an expanded role for products liability; as analysis shifts from drivers' negligence to defects in driving systems, products liability will take a greater slice of the liability pie.[151] Within a products liability framework, plaintiffs could attempt to use consumer expectation and risk-utility tests to prove that automated systems are defective. They will need to demonstrate reasonable alternative designs, which in turn may depend on statistical crash data showing the comparative safety of different automated systems and human-driven vehicles.[152] With respect to the problem of harm caused by emergent AAs, Patrick Hubbard suggests that plaintiffs can try to prove that the "use of emergent behaviour" in a product's design failed to satisfy the standard of a risk-utility test, and that the failure to satisfy this standard caused the harm.[153] This involves demonstrating that, at

147. *Ibid* at § 3.
148. Owen, *supra* note 127 at 984.
149. *Ibid* at 987.
150. Notably, there is an additional threshold issue about whether software operating automated systems—including the data it uses or produces—are products for the purposes of products liability law. See Smith, *supra* note 120 at 45.
151. Smith, *ibid* at 2.
152. *Ibid* at 46.
153. F Patrick Hubbard, "Sophisticated Robots: Balancing Liability, Regulation and Innovation" (2014) 66:5 Fla L Rev 1803 at 1852.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

the time of design and sale, there was an available alternative design that was safer overall.[154] However, Hubbard also admits that emergent characteristics "raise virtually insurmountable proof problems concerning such issues as defectiveness and causation," and "[p]laintiffs who cannot adequately address problems in proving breach of standard and causation will lose their cases."[155] The crucial point is that products liability for design defects is not truly strict, and it does not simply resolve the doctrinal challenges of discovering liability for AA-caused harm. It does not side-step doctrinal constraints like foreseeability and proximate causation.[156] In software and algorithm defect cases, moreover, plaintiffs are likely to face serious technical and financial difficulties: products liability litigation is enormously expensive, requiring expert witnesses for specialized evidence about algorithms and sensor data.[157] It is an imperfect solution to the problem of AA-caused harm.

In sum, there is no straightforward strict products liability remedy for AA-caused harm. Tort liability turns on finding fault in the way AAs are designed, constructed or used, unless it can be shown that deploying AAs is an ultra-hazardous activity subject to strict liability. I have argued that deploying AAs should not be characterized this way. As a result, it appears the costs of AA-caused harm lie where they fall under traditional common law tort analyses.

4. Should AA-Caused Losses Lie Where They Fall As a Matter of Tort Law?

It is not surprising that the costs of AA-caused harm likely lie where they fall under the traditional common law of negligence and strict liability. However, this raises a further question: is this outcome satisfactory as a matter of tort law? Should AA-caused losses should lie where they fall or is it necessary to craft alternative tort standards? In this subsection I consider two alternative accounts pursuant to corrective justice.

The first account suggests that AA-caused loss should lie where it falls (as a matter of tort law), as this doctrinal outcome is indicative of fundamental incompatibility between AAs' functions and tort law. Tort law concerns agent-specific bases of liability, where legal actors wrongfully cause loss to other legal actors. Since AA-caused harm does not implicate wrongful conduct, it falls outside the realm of tort. This does not mean that AA-caused harm is not a matter of moral or political concern. AA-caused harm implicates important issues of distributive justice, an imperative of political morality to ensure the benefits and burdens of AAs are fairly shared by citizens. In this respect, it may be unfair to permit distributors,

154. *Ibid* at 1854-58.
155. *Ibid* at 1851-52.
156. Karnow, *supra* note 3 at 72-74.
157. Jeffrey K Gurney, "Imputing Driverhood: Applying a Reasonable Driver Standard to Accidents Caused by Autonomous Vehicles" in Patrick Lin, Keith Abney & Ryan Jenkins, eds, *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence* (Oxford University Press, 2017) 51 at 59-60.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

owners or users of AAs to externalize accident costs onto victims, while taking a disproportionate share of AAs' benefits. However, the morality of distributive justice does not ground agent-specific reasons to repair misfortunes; it concerns the ongoing participation of all industry contributors, to ensure the burdens of AAs are distributed in proportion to their benefits. Accordingly, the argument goes, since tort is exclusively concerned with agent-specific reasons for liability—organized around normative notions like wrongful conduct and outcome-responsibility—the distribution of costs of AA-caused harm is not the kind of moral issue that tort concerns. Tort is then the wrong institutional forum for its resolution. The construction of liability standards for AA-caused harm falls within legislative territory, where liability regimes can be crafted to realize critical policy goals, such as incentivizing safe and beneficial innovation and compensating victims of AA-caused harm.

A second account suggests, however, that the doctrinal outcome declining liability under common law negligence and strict liability standards is not determinative of corrective justice. It is not determinative of whether there are agent-specific bases of liability for deploying emergent AAs with potential to cause unpredictable harm. Since corrective justice is a normative lens implicit in tort, it provides insight into why AA-caused harm does not straightforwardly ground liability under traditional negligence and strict liability doctrines. This relates to the assumption that the legally relevant social relation for tort liability is an immediate relation between two legal subjects, in which specific correlative rights and duties are given determinate content. As a practice of corrective justice, tort principles interpret interactions to construct legally relevant bilateral units that ground agent-specific bases of liability. The problem of AA-caused harm reflects the problematic nature of AAs' social roles, as mediators of human interaction. AAs are anticipated to do jobs previously undertaken exclusively by humans, such as driving cars and performing surgeries. Certainly, this way of describing AAs is provocative and inexact: AAs are actually sophisticated products. Automated vehicles are sophisticated cars, not drivers, and surgical robots are sophisticated medical tools, not surgeons. Nevertheless, because emergent AAs undertake these jobs with functional independence—not subject to full human control—the normative significance of their effects in the world is arguably not entirely analogous to those of mere products. In the usual case of products, their designs can be traced to human designers, and their uses can be traced to human users. By contrast, with emergent capacities, AAs may act on environmental inputs that are not traceable to particular human designers or users. AAs' outward effects are not necessarily attributable to human agency in the way effects of products typically are. What then is the legal conception of AAs' harmful outward effects? One may argue that they are legally comparable to natural events. Yet, this approach is also unsatisfactory: natural events are not deliberately deployed by humans for beneficial human ends, as AAs are. Arguably, AAs have special social significance, as mediators of human interaction: they are deployed by individuals to perform complex tasks and to interact with other

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

individuals and property. AAs are quasi-social-actors, extraordinary entities without obvious legal classification.

AAs' mediating role adds a layer of complexity to tort relations that is not present in ordinary common law judgments of liability. There is distinct challenge of discovering liability for AA-caused harm because the human (or corporate) supplier or user is simply not the relevant social actor that causes harm; another actor, the AA, was deployed to perform the task instead. If suppliers or users of AAs commit no initial tortious act, they are not exposed to liability. Simply put, suppliers and users of AAs are not in relations with victims of AA-caused harm that are amenable to negligence or strict liability, doctrines that do not capture the relevance of deliberate, though reasonable, deployment of another social actor to perform a task. This does not necessarily mean, however, that deployers of AAs cannot be subject to agent-specific bases of liability, nor does it mean there is no issue of corrective justice for AA-caused harm. It is plausible to argue that the deployment of AAs should, in some cases, ground agent-specific duties of corrective justice in the event of AA-caused harm. AAs are distributed and deployed by suppliers and users with knowledge that they entail an uncontrollable potentiality of harm, at least in isolated cases. AAs' functional independence is a feature of their special utility. Since deployers leverage AAs' efficacy—including their functional independence—to advance their purposes, they should assume some responsibility for AA-caused harm, despite reduced foreseeability and proximate causation. Complete immunity from liability for AA-caused harm is intuitively unsatisfactory, and examination of alternative liability standards is advisable.

V. Alternative Liability Regimes and Reconceiving Tort Law Principles

It is not surprising that some recent commentary reveals a preference to depart from traditional tort principles to regulate AA-caused harm. For example, Kenneth Abraham and Robert Rabin propose a new compensation regime—termed Manufacturers Enterprise Responsibility (MER)—for victims of autonomous vehicles, as an exclusive remedy, replacing victims' rights to sue in tort.[158] MER would operate as a third-party insurance system funded by manufacturers of autonomous vehicles to compensate accident victims without resort to proof of negligence or products liability.[159] Likewise, David Vladeck proposes a form of common enterprise liability: strict liability imposed jointly on the group of manufacturing and designing entities working together to create an AA.[160] As a strict liability standard, it is essentially a court-compelled insurance regime

158. Kenneth S Abraham & Robert L Rabin, "Automated Vehicles and Manufacturer Responsibility for Accidents: A New Legal Regime for a New Era" (2019) 105:1 Va L Rev 127.
159. *Ibid* at 145-56.
160. Vladeck, *supra* note 2 at 129.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

requiring suppliers of AAs to compensate accident victims without regard to notions of fault or defect under negligence or products liability.[161]

Nevertheless, many commentators are less inclined to abandon the tort framework. Patrick Hubbard maintains that tort methods for addressing harm caused by AAs—i.e. products liability governing risks of distribution and negligence governing risks of use—suitably balances goals of innovation and safety, preferable to third party no-fault liability schemes.[162] However, as argued above, and acknowledged by Hubbard, negligence and products liability principles are unlikely to ground liability for harm due to AAs' emergent behaviours. For this reason, alternative lines of commentary consider various analogies to liability for animals,[163] slaves,[164] prisoners[165] and employees.[166] Some commentators are also creatively constructing tort principles specifically for AA-caused harm. Ryan Abbott proposes that as autonomous machines enter into positions previously occupied exclusively by humans, manufacturers' liability for AA-caused harm should be assessed on the basis of the reasonableness of AAs' outward actions, under a negligence standard, rather than investigating potential defects as a matter of products liability.[167] Similarly, in the context of automated driving, Jeffrey Gurney recommends that as operation of vehicles shifts from human drivers to computer systems, the law should treat manufacturers as the drivers of the vehicles subject to a "reasonable driver" standard, avoiding logistical problems and transaction costs of products liability litigation.[168] Displaying equally innovative spirit, Samir Chopra and Laurence White propose an agency framework, viewing AAs as legal agents of human users, grounding vicarious liability for AAs' harmful actions.[169] Several commentators also tentatively present the possibility of granting legal personhood to AAs, linked to self-insured capital funds; in this event, victims could sue AAs directly.[170] Finally, Mark Chinen suggests reframing notions of responsibility for AA-caused harm in associative or collective forms. Under this approach, responsibility is assumed by (groups of) humans associated with AAs, though the AAs themselves may be the only entities causally responsible in any particular instance.[171]

---

161. *Ibid* at 146.
162. Hubbard, *supra* note 153 at 1867-69.
163. Peter M Asaro, "A Body to Kick, but Still No Soul to Damn: Legal Perspectives on Robotics" in Patrick Lin, Keith Abney & George A Bekey, eds, *Robot Ethics: The Ethical and Social Implications of Robots* (MIT Press, 2012) 169 at 176-78.
164. Leon E Wein, "The Responsibility of Intelligent Artifacts: Toward an Automation Jurisprudence" (1992) 6 Harv JL & Tech 103 at 110-11.
165. Chopra & White, *supra* note 59 at 134.
166. *Ibid* at 133.
167. Abbott, *supra* note 18 at 22.
168. Gurney, *supra* note 157 at 59-60.
169. Chopra & White, *supra* note 59 at 127-30 and Mark A Chinen, "The Co-Evolution of Autonomous Machines and Legal Responsibility" (2016) 20:2 Va JL & Tech 338 at 348, where he states: "[a] self-driving car is not dissimilar to a human chauffer," stressing autonomous vehicles' agent-like role akin to employees.
170. Pagallo, *supra* note 53 at 153-54; Chopra & White, *supra* note 59 at 186-89; Vladeck, *supra* note 2 at 150.
171. Chinen, *supra* note 169 at 366, 375-77.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

This emerging literature reveals three general strands of thought regarding the application of tort law to AAs: whether to affirm the applicability of tort, modify it, or abandon it altogether for alternative liability regimes. Accordingly, it directs attention to a set of questions: Is tort law unsuitable for regulating AA-caused harm, accentuating the need to adopt new doctrines or legal regimes emphasizing other bases of liability? Is it possible to analogize AA-caused harm to existing tort categories? Is corrective justice relevant in the context of AA-caused harm?

While the ultimate form of liability regime for AA-caused harm remains unknown, there are two strategies that are worth pursuing. The first strategy is to sidestep tort law—and the morality of corrective justice—for alternative legislative regimes that can spread accident costs fairly, as a matter of enterprise liability, a form of distributive justice. The goal would be to redistribute the burdens of AAs—in this case, their accident costs—in proportion to their benefits.[172] The principled idea is that those who benefit most from AA activity ought to receive a more substantial share of its burden as well. To this end, enterprise liability identifies an accident's responsible cause by ascertaining whether the resulting harm is attributable to a characteristic risk of enterprise, i.e. risk inherent to its ongoing activity.[173] Liability is then the fair price the enterprise pays for reaping benefit from its activity entailing characteristic, though reasonable, risk. In this respect, enterprise liability aims at proportionality of benefits and burdens: if the enterprise benefits from activities that also inherently produce certain burdens, such as its characteristic risk of harm (and resultant harm), it cannot keep the benefits while externalizing the burdens. While designing, manufacturing and selling AAs may be reasonable—and more beneficial than harmful—these activities entail characteristic risk of AA-caused harm due to unpredictable emergent behaviour. Therefore, pursuant to a strict liability standard, it may be fair to require distributors of AAs to internalize AAs' accident costs along with their commercial benefits.

In many cases, however, the ideals of enterprise liability should not be constrained by the institutional structure of tort liability, with its narrow focus on particular interactions between defendants and plaintiffs. Under enterprise liability, the locus of liability is the attribution of accidental harm to characteristic risk inherent to ongoing activity, not to conduct that causes harm in particular interactions. Enterprise liability abstracts from particular interactions, instead focusing on activities' costs over time. It aims to distribute costs of *ongoing* risky activities as the fair price of enterprise. From the viewpoint of enterprise liability, accident-costs are best distributed across all similarly situated enterprise participants who reap benefit from the relevant activity while externalizing its characteristic risk onto others. For this reason, enterprise liability often gains fullest

172. This account of enterprise liability, as a form of distributive justice, is based on Gregory C Keating, "Tort, Rawlsian Fairness and Regime Choice in the Law of Accidents," (2004) 72:5 Fordham L Rev 1857 at 1873 [Keating, "Rawlsian Fairness"].

173. Gregory C Keating, "Strict Liability Wrongs" in John Oberdiek, ed, *Philosophical Foundations of the Law of Torts* (Oxford University Press, 2014) 292 at 306-08.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

expression where losses are spread beyond specific firms, but across entire industries or society wide.[174] In any given activity, the proper extent of loss-spreading schemes ultimately depends on how widespread the benefits of the activity are; after all, the burdens of activities must be distributed in proportion to their benefits. Enterprise liability is, therefore, often best implemented through alternative liability regimes—such as worker's compensation or no-fault auto insurance—rather than judicially, through tort law, where one's share of liability is unjustifiably conditional on factually causing harm in particular interactions.[175] The crucial point is this: if the normative foundation of liability is the redistribution of an activity's risk and costs in proportion to its benefits—so that activity participants who benefit most from the activity imposing characteristic risk also internalize a greater share of its costs—then there is inadequate normative basis to merely shift the accident costs to the actor who causes harm in any particular instance, to the exclusion of other activity participants. All similarly situated activity participants imposing the same characteristic risk should share accident costs to achieve proportional burdens-to-benefits. The implementation of an enterprise liability regime for AA-caused harm should then satisfy two basic requirements: First, its loss-spreading scheme should encompass all similarly-situated activity participants, without singling-out particular defendants who cause (non-negligent) accidental harm in particular instances. Second, its loss-spreading scheme should enable activity participants to internalize accident costs as an ongoing—i.e. forward-looking—cost of enterprise through insurance. Notably, the MER regime proposed by Kenneth Abraham and Robert Rabin falls generally under this rubric, as an explicit alternative to tort law in regulating AA-caused harm.

The second strategy is to reconceive tort principles and AAs' legal classification to suitably assimilate AAs within the tort framework. This demands rigorous thinking about AAs' salient features that should have normative impact on rights and duties of legal persons.[176] In advancing the common law of tort, analogical reasoning is essential: AAs should be compared to, and distinguished from, familiar tort categories. Moreover, as a matter of analogical reasoning, it is valuable to discover an appropriate metaphor that captures the nature of AAs' social impact. Various analogies to tort doctrines—whether liability for products, wild animals or employees—reflect different presumptions about AAs' functional and normative roles within human relations. A suitable metaphor is valuable, as it can guide analysis toward a reasonable doctrinal analogy.

The use of metaphor and analogy plays a large role in legal discourse. Metaphors assist lawyers and lawmakers to distill, and make sense of, complex and unfamiliar phenomena beyond immediate expertise.[177] Where lawyers and

174. Keating, "Rawlsian Fairness", *supra* note 172 at 1906-07.
175. *Ibid* at 1899.
176. See Balkin, *supra* note 20.
177. Meir Dan-Cohen, *Rights, Persons and Organizations: A Legal Theory for Bureaucratic Society*, 2nd ed (Quid Pro Books, 2016) at 38.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

judges "operate on foreign intellectual terrain," metaphors "may fill the epistemic gap between what is known and the knowledge that would be desirable for a sounder decision."[178] Moreover, the use of metaphor is simply natural to doctrinal analysis, a process of reasoning by analogy, aligning new social practices with existing legal concepts and categories. The use of metaphor can also be a method of persuasion, a "means of achieving a rhetorical effect by directly equating disparate concepts," thereby justifying specific legal treatment of an issue.[179] For this same reason, metaphors may be dangerous: if they link disparate and unfamiliar concepts without a "self-limiting or guiding principle," inapt metaphors can be misleading, generating false analogies and "crude approximations of social reality" that lead to improper value judgments and policy prescriptions.[180] Metaphors can also have powerful social effect, potentially shaping individuals' views about important matters without being aware of the influence of metaphor.[181]

A prime instance of law's reliance on metaphor is to understand new technologies, and particularly AAs.[182] AAs are an obscure and esoteric phenomenon, subject to specialized and technical expertise. Since metaphors can shape legal judgments or policies, litigants and judges use metaphors strategically in applying law to new technologies.[183] For example, in early e-mail cases, courts dealt with a battle of metaphors: whether emails are analogous to postcards—which do not entail reasonable expectations of privacy under the Fourth Amendment—or to letters, which receive greater privacy protections.[184] Likewise, consideration of tort liability for AA-caused harm gives rise to a battle of metaphors and analogies, where AAs are compared to products, animals, employees or persons. The use of metaphors for AAs inevitably involves danger of false analogy. It must, then, be done cautiously, in pursuit of a legitimate purpose, and fully aware of any legal fiction it involves.[185] Employing a legal fiction serves a legitimate function if it suitably captures the salient features of AAs that law must account for in determining liability. This is the legitimate role of using a legal metaphor for AAs: to

178. *Ibid*.
179. Calo, "Robots as Legal Metaphors", *supra* note 4 at 211.
180. Dan-Cohen, *supra* note 177 at 38.
181. *Ibid*. For example, where the law bestows legal (metaphorical) personhood upon some entities (e.g. corporations), but not others (e.g. nonhuman animals), it may express distinctions of status and moral value, though perhaps without any such intention. See Notes, "What We Talk About When We Talk About Persons: The Language of a Legal Fiction" (2001) 114:6 Harv L Rev 1745 at 1760.
182. Deborah G Johnson & Mario Verdicchio, "Why Robots Should Not Be Treated Like Animals" (2018) 20:4 Ethics and Information Technology 291 at 292: "it is not surprising that scholars and lay thinkers make use of analogies to think about robots. Analogical reasoning is a common strategy for understanding new phenomena. We use something familiar to understand something unfamiliar or less familiar."
183. Calo, "Robots as Legal Metaphors", *supra* note 4 at 212-13.
184. *Ibid*.
185. LL Fuller, "Legal Fictions" (1931) 25 Ill L Rev 513 at 527-28: "Developing fields of law, fields where new social and business practices are necessitating a reconstruction of legal doctrine, nearly always present 'artificial construction,' and in many cases, outright fictions."

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

conceptualize the normative character of AA-mediated interpersonal relations, facilitating the choice of suitable and commonsensical liability doctrines.

The above discussion reveals a variety of possible legal, and largely metaphorical, conceptions of AAs. Where liability is grounded in products liability, as Patrick Hubbard maintains, the law implicitly views AAs as *things* (products) to be examined for defects. By contrast, Ryan Abbott's suggestion that manufacturers' liability be based on the reasonableness of AAs' behaviours—under negligence—implicitly views AAs as *actors* who are expected to conform to tort's norms of rightful interaction. Similarly, proposals likening liability for AA-caused harm to liability for harm caused by wild animals or employees involve differing conceptions of AAs. The analogy to wild animals implicitly views AAs as *dangerous property* capable of acting independently and erratically (a problematic analogy, as discussed above in section IV). The analogy to employees implicitly views AAs as *legal agents* with capacity to produce legal obligations on behalf of human principals. These distinctions matter; the choice of metaphor involves normative insinuations with doctrinal consequences. Does liability flow from the legal premise that AAs may be deployed only at one's risk, as in the case of keeping wild animals? Or is liability limited to instances where AAs cause tortious harm while acting in the scope of their deployment on behalf of human or corporate principals, resembling vicarious liability for employees' tortious harm? Accordingly, the choice of metaphor and analogy should consider both the conception of AAs implicit in it and its doctrinal outcomes.

The law may need to recognize a new legal category for AAs to properly capture their salient features impacting tort relations. In this respect, Ryan Calo suggests that the law may adopt a "new category of legal subject, halfway between person and object."[186] Similarly, Jack Balkin notes that since AAs substitute for animals or humans in various contexts and for particular purposes, they may be treated as "special-purpose animals" or "special-purpose human beings" depending on the context.[187] According to Balkin, AAs may constitute a new legal category halfway between persons and objects in the sense that their status is partial and contextual, depending on the purposes particular AAs serve.[188] Under this approach, varying metaphors are attributable to various AAs—automated vehicles may be viewed as drivers, whereas robotic pets may be viewed as domesticated animals—resulting in numerous and divergent liability outcomes.

The idea that AAs can constitute a legal category halfway between persons and objects is a valuable framing. Since AAs are expected to act independently in numerous social roles—as artificial medics, drivers, caregivers, etc.—they are not altogether reduceable to *mere* tools. At the same time, AAs are life-less *artifacts*, designed, manufactured and deployed by humans for human purposes. In this sense, AAs constitute an ambiguous middle-ground between full-fledged

186. Calo, "Robotics and Lessons of Cyberlaw", *supra* note 1 at 549.
187. Balkin, *supra* note 20 at 57.
188. *Ibid* at 57-59.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

social actors and mere objects. It may then be sensible to recognize a new legal category to reflect AAs' distinct position: as *pure legal agents* without legal personhood.[189] My contention is that a pure legal agency category can capture the relational structure of AA-mediated human interaction, where AAs—conceived of as functionally independent and rational instrumentalities—are deployed to act exclusively for human purposes. AAs inherently occupy a social role akin to legal agents: they act exclusively *for* their deployers, without purposes of their own. As legal agents, AAs would be empowered to trigger liability consequences for deployers by their own acts. AAs' outward effects would be elevated to the realm of tortious (or non-tortious) *action.* AAs commit torts where their outward behaviours fail to conform to standards of care expected of reasonable persons in the circumstances. Deployers could then be vicariously liable for tortious acts committed by AAs in the course of deployment, circumventing the need to trace AA-caused harm to tortious human agency. This doctrinal outcome is analogous to employers' vicarious liability for torts committed by employees in the course of employment. It is a sensible and pragmatic strategy: liability would be based on the unreasonableness of AAs' outward behaviours, not their internal product defectiveness, avoiding onerous and expensive scrutiny of algorithmic designs for legal defects.

This is effectively the doctrinal strategy recommended by Ryan Abbott and Jeffrey Gurney, though they do not use the language of vicarious liability or legal agency. As stated above, Ryan Abbott proposes that manufacturers' liability for AA-caused harm should be based of the reasonableness of AAs' actions, under a negligence standard. Likewise, Jeffrey Gurney suggests that the law should treat manufacturers as the drivers of autonomous vehicles subject to a "reasonable driver" standard. This is a valuable framing: tort analysis could be directed at evaluating AAs' outward actions, as tortious or non-tortious, without needing to scrutinize AAs' inner algorithmic mechanisms or identify specific human tortious conduct, per se. This approach is essentially a form of vicarious liability, as manufacturers are liable for the tortious acts of AAs they distribute. It reflects their special responsibility for distributing AAs to independently perform specific tasks of deployment, producing characteristic risk (though not real and substantial risk) of tortious harm.

While reconceiving the vicarious liability doctrine is a creative and laudable attempt to assimilate AAs within the tort law framework, it modifies tort law significantly, demanding further study as to whether AAs' pure legal agency status suitably coheres with tort law and corrective justice. The vicarious liability account raises fundamental doctrinal and theoretical questions: Can AAs comprise legal agency relations without legal personhood? Can AAs have, and comply with, fiduciary obligations, which are ordinarily constitutive of legal agency relations? How is the scope of AAs' deployment determined? Can AAs' tortious acts be viewed as frolics, unattributable to the deployment relation? Who are the

189. The idea that AAs can be legal agents without legal personhood is maintained by Chopra & White, *supra* note 59 at 25.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

deployers for the purposes of vicarious liability, AAs' users or suppliers? I will not attempt to provide answers here. It suffices to flag these questions, to highlight some of the doctrinal challenges associated with this account. It also indicates that the doctrinal issue of tort liability for AA-caused harm should be linked to further theoretical inquiry about AAs' position in a broader legal taxonomy.

One may reasonably object to this approach: viewing AAs as legal agents—empowered to trigger liability consequences by their own acts—is an extreme and unprecedented modification of tort law. If possible, the argument goes, it is preferable to adopt more-modest adjustments to tort law for which there is some historical precedent. This may involve reducing causation standards or plaintiffs' burden of proof. It can also involve embracing strict products liability. Such changes would not be unprecedented: a historical-sociological perspective could locate current technological developments within broader legal trends occurring in the early decades of the second half of the 20th century to utilize tort law to mitigate risks or spread accident costs more generally.[190] During that time period, increased use of automobiles, pharmaceuticals and power tools generated public concern about environmental, product and auto safety, leading to wider acceptance of enterprise liability, strict products liability and reduced causation standards in mass-torts cases.[191] These doctrines are prime examples of partial abandonment of traditional tort fault-finding analyses in favour of more impersonal mechanisms for distributing social risk.[192]

However, we should be wary about altering tort law without addressing the problematic issue of fair interpersonal conduct with respect to deploying AAs. Re-establishing strict products liability or reducing causation standards to advance goals of enterprise liability distances tort from its core ideas of wrongful interpersonal conduct and outcome responsibility. To simply reduce the centrality of fault upsets tort's structure of agent-specific responsibility. As noted above, tort's narrow focus on particular interactions between defendants and plaintiffs reduces the fairness of loss-spreading schemes, preventing losses from being shared by all similarly situated enterprise participants benefiting from the relevant activity causing harm. Reduced-fault modifications point toward crafting legislative liability alternatives; they are not well-suited to ground agent-specific duties of repair in tort. Moreover, these reduced-fault modifications react inadequately to the advent of AAs, persisting to treat them as ordinary products, despite their emergent features. The problem of AA-caused harm relates to their awkward fit within tort's basic structure when regarded as mere products. A doctrinal resolution is then best served by thinking carefully about AAs' extraordinary roles as functionally independent and rational instrumentalities, with potential to cause unpredictable harm. The vicarious liability account is valuable for this reason: it provides agent-specific grounds for liability particularly

190. Robert L Rabin, "Tort Law in Transition: Tracing the Patterns of Sociolegal Change" (1988) 23:1 Val U L Rev 1 at 13.
191. *Ibid* at 13-24.
192. *Ibid*.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

appropriate to the AA-deployment relation. If deployers advance their purposes by deploying AAs to perform tasks of deployment, they should assume special responsibility for tortious harm that relates to these tasks, akin to employers' special responsibility for employees' torts that are sufficiently linked to their tasks of employment. In this way, deployers are liable only where AAs' acts are deemed tortious and sufficiently tied to tasks of deployment. It is not strict liability. Deployers' agent-specific duties to repair would then reflect their responsibility for deploying AAs to perform assigned tasks entailing characteristic risk of tortious harm. This is not to say a vicarious liability solution is straightforward; there are various complications, both doctrinal and theoretical, without self-evident solutions, as noted above. Nevertheless, it is the kind of robust solution worth pursuing, one that fundamentally reworks the legal conception of AAs to capture their salient features, guiding plausible tort liability outcomes.

In sum, an adequate legal response to AA-caused harm demands either implementation of legislative liability regimes or modification of tort law itself. Importantly, however, these two paths are not mutually-exclusive; a combination of these approaches is advisable. Tort law solutions to the problem of AA-caused harm may not sufficiently satisfy various policy imperatives with respect to developing AAs: incentivizing innovation, establishing consumer-safety product design standards, deterring cost-inefficient and unsafe product design, reducing or widely-spreading accident costs, compensating victims of AA-caused harm, among others. As matters of policy, these objectives should be implemented through regulatory schemes aimed at fairly distributing the benefits and burdens of AAs. At the same time, there is still good reason to care about tort law. The development of a common law jurisprudence for AA-caused harm is vital if deployment of AAs becomes commonplace. It is certainly conceivable that specific industries—e.g. transportation or healthcare—will be subject to specific regulatory schemes ensuring accident victims receive compensation. The MER regime for autonomous vehicles proposed by Kenneth Abraham and Robert Rabin is one such example. Nevertheless, common law principles of general application are still necessary for any instance of AA-caused harm that is not subject to a specific regulatory scheme. In these instances, courts will have to look to the common law of tort to develop a doctrinal resolution to the problem of AA-caused harm.

Doctrinal analyses of AA-caused harm should certainly not be formalistic. After all, where AAs are analogized to existing tort categories—e.g. products, animals, employees—we should expect divergent and conflicting, yet reasonably logical, doctrinal conclusions. In this respect, recall Judge Cardozo's dictum in *Jacobs & Youngs v Kent*:

> Those who think more of symmetry and logic in the development of legal rules than of practical adaptation to the attainment of a just result will be troubled by a classification where the lines of division are so wavering and blurred.[193]

193. *Jacobs & Youngs, Inc v Kent*, 230 NY 239 at 242-43.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

In hard cases, where simple and mechanistic doctrinal resolutions are unavailable, a sense of justice and good policy should guide choices "between one logic and another."[194] At the same time, judicial formulas designed to achieve a sense of justice must be squared "with principle and with the symmetry of the legal system."[195] Doctrinal analysis is an inductive and reflective process, squaring justice with logic, and logic with justice. While justice guides choices between differing legal precedents, the logic of doctrinal analysis relates these choices to legal method, maintaining law's coherence, consistency and uniform structure.[196] Accordingly, doctrinal analysis of AA-caused harm demands attention to both sound liability outcomes and achieving a "larger and more inclusive unity"[197] within tort law. It should utilize tort categories creatively to capture AAs' social and normative significance, consistent with intuitions about how AAs should impact rights and duties of legal persons.

VI. Conclusion

AA-caused harm generates fresh doctrinal and theoretical challenges for assigning tort liability. Due to machine-learning capabilities, AAs' harmful outward effects may be principally untraceable to tortious actions of designers, manufacturers or users. As a result, pursuant to traditional common law tort doctrines, absent modification, the costs of AA-caused harm likely lie where they fall. Under the law of negligence, AA-caused harm is not easily attributable to tortious conduct because AAs act with functional independence, breaking the chain of proximate causation traceable to legal agency. Moreover, the deployment of AAs should not be subject to strict liability, as it is not exceedingly risky relative to ordinary background risks; in many instances, AAs are projected to produce net safety gains.

This doctrinal outcome relates to tort law's theoretical structure of corrective justice, which resists straightforward application to AA-mediated social relations. As a practice of corrective justice, tort liability applies to bilateral relations constituted by two legal subjects, whereby one actor wrongfully causes loss to the other, grounding an agent-specific duty to repair. Since AA-caused harm is more precisely attributable to the AA itself—which is not a legal subject—its normative impact on rights and duties of legal persons is unclear.

However, permitting costs of AA-caused harm to lie where they fall reflects a problematic incongruity: whereas personally performing tasks causing harm could potentially produce liability, individuals can avoid tort liability altogether by deploying AAs to perform the tasks instead. Straightforward application of negligence and strict liability doctrines—and letting AA-caused losses lie where they fall—is unsatisfactory, as it misconstrues AAs as ordinary products. AAs are

194. Benjamin N Cardozo, *The Nature of the Judicial Process* (Yale University Press, 1921) at 41.
195. *Ibid* at 39.
196. *Ibid* at 41, 46.
197. *Ibid* at 46.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press

functionally independent and rational instrumentalities deployed to perform specific tasks. Since deployers utilize AAs' functional independence to advance their purposes, they should assume some responsibility for AA-caused harm. To this end, creative modification or reconception of liability principles is advisable. A successful doctrinal resolution is served by thinking carefully about AAs' salient features, including their roles as functionally independent instrumentalities, to determine their proper normative impact on interpersonal rights and duties, guiding sound liability outcomes. One promising solution is to reconceive the doctrine of vicarious liability to make deployers liable for tortious harm committed by AAs in the course of deployment. Under this approach, AAs would resemble legal agents, empowered to trigger liability consequences for deployers by their own acts, akin to employees. The vicarious liability approach admittedly alters the application of tort law, and its doctrinal elements have not been sufficiently worked out in this essay. Nevertheless, it is an initial offering to reconceptualize AAs to reflect their emergent features and to ground agent-specific reasons for liability, consistent with tort's basic structure.

https://doi.org/10.1017/cjlj.2020.3 Published online by Cambridge University Press