Texas A&M Law Review

Volume 4 | Issue 2

9-16-2017

The Trolley and the Pinto: Cost-Benefit Analysis in Automated Driving and Other Cyber-Physical Systems

Bryant Walker Smith

THE TROLLEY AND THE PINTO: COST-BENEFIT ANALYSIS IN AUTOMATED DRIVING AND OTHER CYBER-PHYSICAL SYSTEMS

*by Bryant Walker Smith**

TABLE OF CONTENTS

I. INTRODUCTION

Automated driving has attracted substantial public and scholarly attention. This brief Article describes how that attention has brought new fame to a classic philosophical thought experiment (the "trolley problem"), critiques how this thought experiment has been applied in that context, proposes a more practical extension of that experiment based on risk rather than harm, notes that this extension may still involve programming value judgments, argues with reference to the Ford Pinto debacle that these judgments could inflame juries or the public at large, and emphasizes the need for appropriately focused public discussion of these issues. The Article may be especially relevant to developers and regulators of cyber-physical systems,[1] in-

* Assistant Professor of Law and (by courtesy) Engineering, University of South Carolina; Affiliate Scholar, Center for Internet and Society at Stanford Law School; Adjunct Clinical Professor of Law, University of Michigan Law School; Member, Federal Advisory Committee on Automation in Transportation; Chair, Committee on Emerging Technology Law of the Transportation Research Board of the National Academies; Reporter, Study Committee on State Regulation of Driverless Cars; Chair, Planning Task Force of the On-Road Automated Driving Standards Committee of SAE International. This Article reflects my own views rather than those of the organizations with which I am affiliated. I am grateful to Kelly Meyers, my high school principal, for her selfless dedication to my education. I would also like to thank Edwin Olson, Richard Bryant, and the editors of the Texas A&M Law Review. My publications are available at newlypossible.org.

1. Cyber-physical systems are "engineered systems that are built from, and depend upon, the seamless integration of computational algorithms and physical components." *Cyber-Physical Systems (CPS)*, NAT'L SCI. FOUND., https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503286 [https://perma.cc/2LHM-R5BJ]; *see also Cyber-Physical Systems*, NIST, https://www.nist.gov/el/cyber-physical-systems [https://perma.cc/4TQJ-XK5K].

DOI: https://doi.org/10.37419/LR.V4.I2.2

cluding the automated driving systems[2] that operate self-driving vehicles.

II. The Trolley Problem

The trolley problem is a philosophical thought experiment that posits a runaway streetcar about to strike a group of people standing on the trolley track.[3] If the trolley is diverted to an alternate track, it will strike only one person. A track worker, who can easily flip the track switch, must choose between causing the deaths of several people through inaction and causing the death of one person through action. The numerous variations[4] of this scenario provide concrete and constrained facts through which philosophers can explore specific value judgments within specific ethical frameworks.[5]

As automated driving has captured the public imagination, self-driving vehicles have replaced trolleys in many of these scenarios. In 2012, I suggested that my readers:

> Imagine you are driving down a narrow mountain road between two big trucks. Suddenly, the brakes on the truck behind you fail, and it rapidly gains speed. If you stay in your lane, you will be crushed between the trucks. If you veer to the right, you will go off a cliff. If you veer to the left, you will strike a motorcyclist. What do you do? In short, who dies?[6]

In 2013, Patrick Lin imagined:

> On a narrow road, your robotic car detects an imminent head-on crash with a non-robotic vehicle — a school bus full of kids, or perhaps a carload of teenagers bent on playing "chicken" with you, knowing that your car is programmed to avoid crashes. Your car,

2. An automated driving system is the combination of hardware and software that operates a self-driving vehicle. *See Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, SAE Int'l, http://standards.sae.org/j3016_201609/ [https://perma.cc/E8U9-M29Y].

3. *See* Judith Jarvis Thomson, Comment, *The Trolley Problem*, 94 Yale L.J. 1395, 1414–15 (1985).

4. *Id.*; *see also* John Mikhail, *Moral Grammar and Intuitive Jurisprudence: A Formal Model of Unconscious Moral and Legal Knowledge*, *in* 50 The Psychology of Learning and Motivation: Moral Judgment and Decision Making 27, 32–35 (Daniel M. Bartels et al. eds., 2009); F.E. Guerra-Pujol, *Trolley Problems*, 63 Drake L. Rev. Discourse 101, 119 (2014); Michael S. Pardo & Dennis Patterson, *Philosophical Foundations of Law and Neuroscience*, 2010 U. Ill. L. Rev. 1211, 1239–40 (2010).

5. *See, e.g.*, Mikhail, *supra* note 4, at 31 ("Despite their obvious limitations, trolley problems are a useful heuristic . . . and their artificiality is a virtue, not a vice, in this regard. These hypothetical cases must be supplemented with more realistic probes drawn from other branches of law, policy, and everyday life, however, if moral competence is to be adequately understood.").

6. Bryant Walker Smith, *Driving at Perfection*, Stan. L. Sch. Ctr. for Internet & Soc'y: Blog (Mar. 11, 2012, 3:20 PM), https://cyberlaw.stanford.edu/blog/2012/03/driving-perfection [https://perma.cc/W3QV-DC7Q].

> naturally, swerves to avoid the crash, sending it into a ditch or a tree and killing you in the process.[7]

In 2014, Noah J. Goodall described:

> An automated vehicle . . . traveling on a two-lane bridge when a bus that is traveling in the opposite direction suddenly veers into its lane . . . . The automated vehicle must decide how to react with the use of whatever logic has been programmed in advance. The three alternatives are as follows: 1. Veer left and off the bridge, which guarantees a severe, one-vehicle crash; 2. Crash head-on into the bus, which will result in a moderate, two-vehicle crash; and 3. Attempt to squeeze [past] the bus on the right.[8]

In 2015, Jeffrey K. Gurney highlighted six common hypotheticals, including one in which:

> An autonomous vehicle encounters a situation in which it must strike one of two motorcyclists. To the vehicle's front-left is a motorcyclist who is wearing a helmet. To the vehicle's front-right is a motorcyclist who is not wearing a helmet. Which motorcyclist should the autonomous vehicle strike?[9]

In 2016, Jean-François Bonnefon, Azim Shariff, and Iyad Rahwan asked study participants:

> [T]o indicate how likely they would be to buy an [automated vehicle] programmed to minimize casualties (which would, in these circumstances, sacrifice them and their co-rider family member), as well as how likely they would be to buy [such a vehicle] programmed to prioritize protecting its passengers, even if it meant killing 10 or 20 pedestrians.[10]

Other examples abound,[11] and 2017 will likely bring many more.

These examples tend to move the trolley problem from a thought experiment for philosophers to a practical challenge for programmers and policymakers. Gurney wrote about the "serious ethical and legal questions" raised by the trolley problem.[12] Lin characterized similar questions as "the most urgent area of . . . programming" for automated driving.[13]

---

7. Patrick Lin, *The Ethics of Saving Lives with Autonomous Cars Is Far Murkier than You Think*, WIRED (July 30, 2013, 6:30 AM), https://www.wired.com/2013/07/the-surprising-ethics-of-robot-cars [https://perma.cc/2UJE-LVDD].

8. Noah J. Goodall, *Ethical Decision Making During Automated Vehicle Crashes*, 2424 TRANSP. RES. REC. 58, 60 (2014).

9. Jeffrey K. Gurney, *Crashing into the Unknown: An Examination of Crash-Optimization Algorithms Through the Two Lanes of Ethics and Law*, 79 ALB. L. REV. 183, 197 (2016).

10. Jean-François Bonnefon et al., *The Social Dilemma of Autonomous Vehicles*, 352 SCI. 1573, 1574 (2016).

11. For more, see Gurney, *supra* note 9.

12. *See* Gurney, *supra* note 9, at 186.

13. Patrick Lin, *Why Ethics Matters for Autonomous Cars*, *in* AUTONOMES FAHREN 69, 69 (Markus Maurer et al. eds., 2015).

Variations of the trolley problem could conceivably arise in the real world of driving. That world, after all, is one in which more than a billion motor vehicles[14] travel trillions of miles every year,[15] giving rise to myriad situations that are tragic, absurd, or perplexing.[16] Advanced automotive technologies, including driving automation systems and automated emergency intervention systems,[17] will eventually confront these situations. Cyber-physical systems more generally—including those on the ground, in the air, in the home, and in the body—could also face similar dilemmas.

On the road, choices about how to respond to these dilemmas could accordingly shift from reactive drivers to proactive designers. The hypothetical conductor from the classic trolley problem has the time, knowledge, and—at least within the problem's constrained world—power sufficient to fully contemplate all possible courses of action and to then implement one that is ethically preferable to the others. The actual human drivers who could conceivably face analogous dilemmas do not enjoy these luxuries, but driving automation systems (and their designers) might.

III. The Problem with the Trolley Problem

In the last few years these crash dilemmas have received significant public attention.[18] This popular preoccupation has created the expectation that every conceivable ethical quandary must be identified and satisfactorily resolved before an automated system should or even can be deployed. It also obscures the strongest rebuttal to that expectation by distracting from the substantial risks of today's driving. In other words:

> The fundamental ethical question . . . is this: In the United States alone, tens of thousands of people die in motor vehicle crashes every year, and many more are injured. Automated [driving sys-

14. *See Registered Vehicles Data by Country*, World Health Org., http://apps.who.int/gho/data/node.main.A995 [https://perma.cc/FR6K-5CRX] (last updated Nov. 30, 2015).

15. Bureau of Transp. Statistics, *Table 1-36: Roadway Vehicle-Miles Traveled (VMT) and VMT Per Lane-Mile by Functional Class(a)*, U.S. Dep't of Transp., https://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_statistics/html/table_01_36.html [https://perma.cc/VQ3Z-2QN2].

16. *See, e.g.*, Raphael Orlove, *Russian Driver Evading Cops Through an Airport Terminal Is Some Jason Bourne Shit*, JALOPNIK (Dec. 23, 2016, 9:40 AM), http://jalopnik.com/tag/russian-dash-cams [https://perma.cc/Y38Z-NY9Q].

17. *See, e.g.*, Bryant Walker Smith, *How Governments Can Promote Automated Driving*, N.M. L. Rev. (forthcoming 2016), https://ssrn.com/abstract=2749375; Bryant Walker Smith, *Slow Down that Runaway Ethical Trolley*, Stan. L. Sch. Ctr. for Internet & Soc'y: Blog (Jan. 12, 2015, 3:42 PM), https://cyberlaw.stanford.edu/blog/2015/01/slow-down-runaway-ethical-trolley [https://perma.cc/92DA-C3D3] [hereinafter *Runaway Ethical Trolley*].

18. Smith, *Runaway Ethical Trolley*, *supra* note 17; *cf.* Bryant Walker Smith, *Automated Driving and Product Liability*, Mich. St. L. Rev. (forthcoming 2017), https://papers.ssrn.com/sol3/papers2.cfm?abstract_id=2923240.

> tems] have great potential to one day reduce this toll, but the path to this point will involve mistakes and crashes and fatalities. Given this stark choice, what is the proper balance between caution and urgency in bringing these systems to the market? How safe is safe enough?[19]

This question is less concrete than a runaway trolley, but its impact on lives saved or lost is much greater. Regardless, even setting aside this larger concern, superficial discussion of the trolley problem tends to miss three key points.

First, even if they are eventually more capable than humans, automated systems will be neither omniscient nor omnipotent:

> Automation does not mean an end to uncertainty. How is an automated vehicle (or its designers or users) to immediately know what another driver will do? How is it to precisely ascertain the number or condition of passengers in adjacent vehicles? How is it to accurately predict the harm that will follow from a particular course of action? Even if specific ethical choices are made prospectively, this continuing uncertainty could frustrate their implementation.[20]

Second, by assuming this omniscience, the trolley problem unhelpfully narrows the discussion from risk to just harm. "The risk of a particular harm is the product of the probability of that harm and the severity of that harm; the risk of an act or omission is the sum of the risks of the particular associated harms."[21] The trolley problem focuses only on harms to the exclusion of probabilities. To their credit, some authors do recognize this distortion. Goodall, for example, noted:

> These [crash] outcomes . . . are not certain. The automated vehicle's path-planning algorithm would have to determine quickly the range of possible outcomes for each considered path, the likelihood of those outcomes occurring, and the algorithm's confidence in these estimates on the basis of the quality of sensor data and other factors.[22]

But third, even when accounting for probability, the trolley problem addresses only crash decisions rather than the driving decisions that can lead to those crashes. Drivers rarely choose which of two obstacles to hit, but they constantly decide how fast to travel and how aggressively to behave vis-à-vis other road users. Similarly, automated driving systems and other complex cyber-physical systems will necessarily make decisions about what risks to accept long before any physical harm is inevitable or even probable.

---

19. Smith, *Runaway Ethical Trolley*, *supra* note 17.

20. *Id.*

21. Bryant Walker Smith, *Lawyers and Engineers Should Speak the Same Robot Language*, *in* ROBOT LAW 78, 85 (Ryan Calo et al. eds., 2016).

22. *See* Goodall, *supra* note 8, at 61.

IV. Beyond the Trolley Problem

Existing in the physical world necessarily entails risk, and the larger question about what risks to accept is much broader than the domain of unavoidable crashes. This is especially true when a cyber-physical system makes decisions based on predictions or assumptions about other actors or about the environment in which they act.

Consider, for example, an automated vehicle merging onto a freeway in dense traffic. There may be a gap in that traffic sufficient for merging—but only if every other vehicle maintains its current lane and speed. Alternately, a sufficient gap may emerge—but only if another vehicle slows slightly to allow the automated vehicle to enter. These two simplified cases each depend on the behavior of other motorists. The first relies on no change in that behavior, whereas the second relies on such a change.

Advanced cyber-physical systems may predict these behaviors. A Google patent cited by Goodall describes a similar situation in which an automated vehicle is traveling in the center lane of a freeway while another vehicle is overtaking it in the left-hand lane.[23] If that left-hand lane is clear, then the automated driving system "could determine that it is highly likely that the other vehicle will continue to overtake [the automated vehicle] and remain in the left-hand lane. Thus, the confidence level corresponding to [this] predicted behavior . . . could be high, such as 90%."[24] In contrast, if that left-hand lane is blocked by a third vehicle, then the automated driving system could determine that there is only a 50% chance of the overtaking vehicle remaining in its lane (while slowing) and a 50% chance that it may instead cut in front of the automated vehicle.[25]

An actual crash involving one of Google's test vehicles illuminates the role that prediction plays in navigation. After moving to one side of a wide travel lane in anticipation of a right turn, the test vehicle encountered sand bags that were obstructing its path.[26] As the vehicle "began to proceed back into the center of the lane to pass the sand bags," it struck a public bus that "was approaching from behind."[27] Both the test vehicle and its safety driver had detected the bus but

23. Modifying Behavior of Autonomous Vehicle Based on Predicted Behavior of Other Vehicles, U.S. Patent No. 8,655,537 col. 8 l. 51–64 (filed May 23, 2013) (issued Feb. 18, 2014).

24. *Id.* at col. 9 l. 3–11.

25. *Id.* at col. 9 l. 12–20.

26. State of Cal. Dep't of Motor Vehicles, Report of Traffic Accident Involving an Autonomous Vehicle (Feb. 23, 2016), https://www.dmv.ca.gov/portal/wcm/connect/3946fbb8-e04e-4d52-8f80-b33948df34b2/Google_021416.pdf?MOD=AJPERES [https://perma.cc/ZL4S-QG7P].

27. *Id.*

"predicted that it would yield."[28] Google subsequently updated its software so that its "cars will more deeply understand that buses and other large vehicles are less likely to yield . . . than other types of vehicles."[29]

These examples speak to probability, but they do not explicitly address harm. In this way, they mirror the trolley problem's focus on harm to the exclusion of probability. But so long as either the magnitude of potential harm is nonzero or the probability of that harm is nonzero, then both components of risk will matter for advanced cyber-physical systems.

In other words, these cyber-physical systems will need to balance risks rather than merely harms or probabilities in isolation. If an automated vehicle will not merge onto a freeway unless there is no conceivable scenario in which harm may occur, then it will not merge at all. But not all potential harms are created equal: A rapid deceleration to avoid a crash may be preferable to a low-speed rear-end crash, which itself may be preferable to a high-speed angular crash.

This balancing is basic cost-benefit analysis, where cost is equivalent to risk. In the example of the merging vehicle, the costs of not merging can be compared with the costs of merging. This comparison could happen in several ways that fall along a spectrum from implicit to explicit. A designer might specify the threshold of confidence in a non-adverse outcome required for a cyber-physical system to undertake a potentially harmful action. The designer might instead instruct the cyber-physical system to compare the respective costs of potential actions based on specific parameters. Or the cyber-physical system itself might be tasked with achieving a broader objective by identifying and then optimizing these parameters based on a massive set of actual or simulated data.

These costs may have different units (if any), including time, money, energy, discomfort, injury, and death, to name a few. For example, stopping on a freeway onramp rather than merging aggressively may increase both travel time and the chance of a rear-end collision but reduce the chance of an angular collision. Similarly, traveling five miles under the speed limit on a city street may increase travel time but reduce both the chance of a collision with a pedestrian and the severity of the pedestrian's injuries should such a collision nonetheless occur.[30] A cyber-physical system tasked with meta-optimization may even identify tradeoffs or behaviors without an obvious relationship to safety.

---

28. GOOGLE, GOOGLE SELF-DRIVING CAR PROJECT MONTHLY REPORT (Feb. 2016), https://static.googleusercontent.com/media/www.google.com/en//selfdrivingcar/files/reports/report-0216.pdf [https://perma.cc/Z2F4-CNHN].

29. *Id.*

30. Bryant Walker Smith, *Regulation and the Risk of Inaction*, *in* AUTONOMES FAHREN 593, 605 (Markus Maurer et al. eds., 2015).

Regardless of how cost-benefit analysis is ultimately performed—whether by specifying confidence thresholds, converting harms to unitless cost factors for use by the cyber-physical system, or defining the meta criteria with which the system itself optimizes these cost factors—such analysis necessarily involves value judgments.

V. The Pinto Problem

Value judgments about the relationship between money and safety are already fraught. As Mark Geistfeld wrote:

> Of course, no one really wants to spend everything on safety. But the widespread resistance to tradeoffs between safety and money is plausibly linked to the principle that "safety matters more than money," what I refer to as the safety principle. The principle has been embraced by many moral philosophers and is reflected in important legal practices.[31]

The Ford Pinto debacle in the 1970s stands as a prominent example of this safety principle.[32] To compete with imported subcompacts, Ford Motor Company rapidly conceived the Pinto, which was intended to weigh no more than 2,000 pounds and to cost no more than $2,000.[33] At the same time, Ford became aware of hazards related to the design of some of its fuel tanks, including those on the Pinto.[34] Although Ford engineers identified several potential fixes, the company adhered to the original design.[35]

Around the same time, Ford prepared an internal memorandum that, while neither specific to the Pinto nor focused on the rear-end collisions in which the Pinto's fuel tanks were particularly susceptible, did explicitly document the kind of cost-benefit analysis used both by the company and by the National Highway Traffic Safety Administration (NHTSA).[36] This "infamous memo"[37] concluded that the benefits of preventing certain fuel tank fires did not justify the costs of making

31. Mark Geistfeld, *Reconciling Cost-Benefit Analysis with the Principle that Safety Matters More than Money*, 76 N.Y.U. L. Rev. 114, 116–17 (2001).

32. *Id. See generally* Mark Dowie, *Pinto Madness*, Mother Jones, Sept.–Oct. 1977, at 18; John R. Danley, *Polishing up the Pinto: Legal Liability, Moral Blame, and Risk*, 15 Bus. Ethics Q. 205 (2005); Dennis A. Gioia, *Pinto Fires and Personal Ethics: A Script Analysis of Missed Opportunities*, 11 J. Bus. Ethics 379 (1992); Gary T. Schwartz, *The Myth of the Ford Pinto Case*, 43 Rutgers L. Rev. 1013, 1013–47 (1991); Anthony J. Sebok, *Punitive Damages: From Myth to Theory*, 92 Iowa L. Rev. 957, 1036 (2007); Stuart M. Gordon & Diane R. Crowley, *Defending the Punitive Damages Claim*, 49 Ins. Couns. J. 300 (1982); Paul J. Zwier, *Due Process and Punitive Damages*, 1991 Utah L. Rev. 407 (1991).

33. Gioia, *supra* note 32, at 380. For a classic Pinto ad, see Thoroughbred Ford, *Old Ford Pinto Commercial*, YouTube (Apr. 16, 2008), https://www.youtube.com/watch?v=8Zn56BUdi0o.

34. Gioia, *supra* note 32, at 380.

35. *Id.* at 381.

36. *Id.*

37. *Id.* at 381–83.

the requisite improvements to fuel tanks.[38] The benefits of $49.5 million were based on an estimate of 180 burn deaths at $200,000 per death (consistent with NHTSA's own valuation of statistical life),[39] 180 serious burn injuries at $67,000 per injury, and 2,100 burned vehicles at $700 per vehicle.[40] The costs of $137 million were based on an additional expense of $11 per vehicle for 12.5 million vehicles.[41]

By the late 1970s, stories of exploding Pintos were spreading. *Pinto Madness*, a scathing magazine article that would later receive both acclaim[42] and critique,[43] depicted Ford as deliberately sacrificing safety for profit.[44] Describing Ford's internal memo, the article pointedly shamed the company for using cost-benefit analysis:

> Ford had gotten the federal regulators to agree to talk auto safety in terms of cost-benefit analysis. . . . Furnished with this useful tool, Ford immediately went to work using it to prove why various safety improvements were too expensive to make. Nowhere did the company argue harder that it should make no changes than in the area of rupture-prone fuel tanks.[45]

Pinto Madness also described how one fuel tank fire had killed a Pinto's driver and severely burned her 13-year old passenger.[46] An outraged jury in this case, *Grimshaw v. Ford Motor Co.*, subsequently returned a $125 million punitive damages award against Ford, though

38. *Id.* at 381.

39. *Id.* Applying a 2017 dollar valuation, the $200,000 figure is equivalent to about $1.1 million (referencing the consumer price index). *See* Lawrence H. Officer & Samuel H. Williamson, *The Annual Consumer Price Index for the United States, 1774-2015*, MEASURINGWORTH, http://www.measuringworth.com/uscpi/ (last visited Feb. 3, 2017). Recent cost-benefit analyses tend to use estimates in the upper-seven figures. *See* CURTIS W. COPELAND, CONG. RES. SERV., R41140, HOW AGENCIES MONETIZE "STATISTICAL LIVES" EXPECTED TO BE SAVED BY REGULATIONS (2010); *see also* U.S. DEP'T OF TRANSP., GUIDANCE ON TREATMENT OF THE ECONOMIC VALUE OF A STATISTICAL LIFE (VSL) IN U.S. DEPARTMENT OF TRANSPORTATION ANALYSES (2013), https://www.transportation.gov/sites/dot.dev/files/docs/DOT%202013%20Signed%20VSL%20Memo.pdf [https://perma.cc/67TS-NCEW] [hereinafter GUIDANCE ON TREATMENT OF THE ECONOMIC VALUE OF A STATISTICAL LIFE].

40. Gioia, *supra* note 32, at 381.

41. *Id.*

42. In 1977, *Pinto Madness* received "a citation for 'distinguished service in journalism' from Sigma Delta Chi, the national journalists' fraternity, and the National Magazine Award in the 'Public Service' category." *See* Adam Hochschild, *Fear and Winning in San Francisco*, MOTHER JONES, July 1978, at 3. Although many sources have reported that this article earned a Pulitzer Prize, *see, e.g.*, GOOGLE, https://www.google.com (search "Pinto Madness Pulitzer Prize" and view the various results claiming the article won a Pulitzer), the organization's website has no record of such an award to the author or for the article, *see Prize Winners by Year*, PULITZER PRIZES, http://www.pulitzer.org/prize-winners-by-year (last visited Feb. 16, 2017). However, the National Magazine Awards have been described as "the equivalent of the newspaper world's Pulitzer Prizes." Hochschild, *supra*.

43. *See* Schwartz, *supra* note 32, at 1017.

44. Dowie, *supra* note 32, at 18, 20.

45. *Id.* at 24.

46. *Id.* at 18 (fictionalizing the name).

this would later be reduced to $3.5 million.[47] Although this jury was not presented with Ford's cost-benefit analysis, it did see a separate Ford engineering analysis that described both a $5 fix and a $10 fix to the Pinto's fuel tank.[48]

In the aftermath of *Pinto Madness* and *Grimshaw*, Ford recalled its 1971–1976 Pintos, quietly settled most of its remaining Pinto-related lawsuits, and defended itself—ultimately successfully—against three charges of reckless homicide brought by the state of Indiana.[49] 1980 was the Pinto's last model year.[50]

The Pinto is far from the only instance in which cost-benefit analysis likely drove safety-relevant design decisions. Indeed, a later case involving another internal memo from the 1970s—this one by a General Motors engineer who "estimated that it would cost $2.40 per car to settle lawsuits resulting from any deaths, as compared with $8.59 to fix the fuel-tank problem"[51]—produced an even greater punitive damage award.[52] This award, in *Anderson v. General Motors Corp.*, also prompted *The Economist* to lament that "the message of the GM award is that cost-benefit analyses, particularly on safety, should not be carried out and in any event should never be written down."[53]

Future litigation may reveal whether companies still explicitly conduct and document cost-benefit analyses for safety-relevant decisions. (NHTSA does.)[54] Even where such an analysis would support an ex-

47. *Id.* at 208.

48. *See, e.g.*, Grimshaw v. Ford Motor Co., 119 Cal. Rptr. 348, 370–71 (Cal. Ct. App. 1981); Schwartz, *supra* note 32, at 1057–60; Michael A. Schmitt, *A Meaning of Corporate Social Responsibility—the Jury Is in*, 8 DEL. J. CORP. L. 211, 231 (1983).

49. Paul J. Becker, Arthur J. Jipson, & Alan S. Bruce, State of Indiana v. Ford Motor Company *Revisited*, 26 AM. J. CRIM. JUST. 181, 181–82 (2002).

50. Robert Sherefkin, *Lee Iacocca's Pinto: A Fiery Failure*, AUTOMOTIVE NEWS (June 16, 2003, 12:01 AM), http://www.autonews.com/article/20030616/SUB/306160770/lee-iacoccas-pinto%3A-a-fiery-failure [https://perma.cc/8BEJ-HKDQ].

51. *GM and the Law*, ECONOMIST (July 15, 1999), http://www.economist.com/node/222948 [https://perma.cc/9RV6-2MG7].

52. *See* Anderson v. Gen. Motors Corp., No. BC116926, 1999 WL 34868875, at *2 (Cal. Super. Ct. Aug. 26, 1999); *see also* Peter Y. Hong, *Judge Cuts Award Against GM to $1.2 Billion*, L.A. TIMES (Aug. 27, 1999), http://articles.latimes.com/1999/aug/27/local/me-4217 [https://perma.cc/Y2BR-723D].

53. *GM and the Law*, *supra* note 51, *quoted in* Geistfeld, *supra* note 31, at 116.

54. NHTSA's federal motor vehicle safety standards are still based on cost-benefit analyses that explicitly account for the value of a statistical life. *See, e.g.*, GUIDANCE ON TREATMENT OF THE ECONOMIC VALUE OF A STATISTICAL LIFE, *supra* note 39. An international functional safety standard for automotive electronics (ISO 26262) suggests a more subjective approach. *See* INT'L ORG. FOR STANDARDIZATION, ISO 26262-9:2011 ROAD VEHICLES — FUNCTIONAL SAFETY — PART 9: AUTOMOTIVE SAFETY INTEGRITY LEVEL (ASIL)-ORIENTED AND SAFETY-ORIENTED ANALYSES, http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=51365 [https://perma.cc/LVA2-95XW]. Pursuant to this standard, a hazard is classified into one of five automotive safety integrity levels (ASILs) based on the likelihood of exposure to that hazard, the driver's ability to prevent injury in the case of that exposure, and the severity of injury if it is not prevented. Each ASIL is then associated with different safety measures.

isting design over a proffered alternative, defendants in product liability cases may nonetheless wish to argue in terms of other tradeoffs, including product functionality, customer convenience, and consumer choice. At some point, however, a company's cost-benefit analysis may be discoverable not only in its memos but also in (or at least through) its code.

In product liability litigation involving an advanced cyber-physical system, a plaintiff might point to the system's computer code or training data as evidence of the defendant's breach or of the product's defect. Inherent in this code could be certain value judgments that would be subject to critique. These judgments could create a perception problem for defendants similar to that presented by cost-benefit analysis: clear documentation that companies had prioritized something other than safety.

For example, that code might specify a confidence threshold or a cost factor for a decision that had resulted in harm. A 90% confidence threshold (to use the example from the patent discussed earlier) might be represented as acceptance of a 10% chance of harm. A cost factor for a collision that is weighted 10,000 times greater than the corresponding cost factor for travel delay might be represented as a willingness to sacrifice lives for speed.

Designers might be asked to justify each of those valuations in front of (or at least in anticipation of) juries that could construe concrete explanations as cold—but abstract explanations as sloppy. Cost-benefit analysis that values lives can offend precisely because it is so concrete, and yet failing to substantiate an input or calibration might suggest indifference of another kind.

Claims involving cyber-physical systems may be complex, and plaintiffs will likely face their own challenges. However, because the code that creates such a system will explicitly or implicitly contain fundamental value judgments, this code could be as potent—for better or worse—as the memos in *Grimshaw* and *Anderson*. And it may be as public as the one in *Pinto Madness*.

VI. Conclusion

Prospective public discussion of the value judgments inherent in cyber-physical systems is important.[55] Many of these judgments, however, are far more pertinent and practical than the stylized choice at the heart of the classic trolley problem. They include, among others,

55. Lin, *supra* note 13, at 82 ("The larger challenge, though, isn't just about thinking through ethical dilemmas. It's also about setting accurate expectations with users and the general public who might find themselves surprised in bad ways by autonomous cars; and expectations matter for market acceptance and adoption."); Smith, *supra* note 30, at 601 ("[E]ncouraging companies to disclose information necessary to their safety case . . . could help educate regulators and the broader public about the capabilities and limitations of these emerging technologies.").

balancing the risks of the status quo with the risks of innovation, trading off between safety and utility, and assigning an economic value to a statistical life.

The U.S. Department of Transportation is an agency that has encouraged this discussion. In its Automated Vehicles Policy Guidance, NHTSA envisions that developers of automated driving systems will perform and document a safety assessment that covers fifteen key points, including system safety, crashworthiness, privacy, and ethical considerations.[56] This safety assessment evokes my own proposal for public safety cases in which developers of automated driving systems (and other cyber-physical systems with the potential to cause substantial harm) publicly and concretely describe their safety philosophy for the entire lifetime of their systems.[57] That proposal in turn draws from the flexible and targeted approach the Federal Aviation Administration (FAA) takes toward new aviation systems.[58]

This public discussion can also provide a shared basis for key inputs into these cyber-physical systems. By sharing data, developers could more accurately and consistently assess probabilities across their various platforms.[59] And by conclusively establishing and updating relevant cost factors (including the value of statistical life),[60] federal agencies could provide a more credible basis for these developers to assess harms. Together, substantiation of probabilities and harms can inform regulators, developers, and—critically—the public at large in understanding inevitable risk and navigating the values that it implicates.

---

56. U.S. DEP'T OF TRANSP., FEDERAL AUTOMATED VEHICLES POLICY 11–36 (2016), https://www.transportation.gov/sites/dot.gov/files/docs/AV%20policy%20guidance%20PDF.pdf [https://perma.cc/RRN3-SS9L].

57. *See* Smith, *supra* note 30; *see also* Smith, *supra* note 17.

58. *See* 14 C.F.R. § 11.81(e) (2011) (requiring that a petitioner for an exemption to an FAA rule provide "[t]he reasons why granting the exemption would not adversely affect safety, or how the exemption would provide a level of safety at least equal to that provided by the rule from which [the petitioner] seek[s] the exemption").

59. NHTSA encourages this. *See* U.S. DEP'T OF TRANSP., FEDERAL AUTOMATED VEHICLES POLICY, *supra* note 56, at 17–19, 99 ("[NHTSA] will explore a mechanism to facilitate anonymous data sharing among those parties testing and deploying HAVs. The mechanism will facilitate sharing that complies with antitrust and competition law requirements, perhaps by using a third-party aggregator. While the specific data elements to be shared will need further refinement, the mechanisms for sharing can be established.").

60. *See supra* note 39. In one of my courses, my students and I spend two hours discussing this value. *See* Bryant Walker Smith, *Seminar on the Law of the Newly Possible*, NEWLYPOSSIBLE.ORG, https://newlypossible.org/wiki/index.php?titleseminar_on_the_Law_of_the_Newly_Possible#Valuing_Life [https://perma.cc/HA7Q-N3HQ] (last modified Feb. 6, 2016). Before they read about established methods for determining this number, they first come up with their own estimate. My students have approximated several techniques, including willingness to pay, willingness to accept, and tort law's approach to damages. Regardless of the method, however, with few exceptions they each arrive at a number between $2 million and $20 million.