California Western Law Review

Volume 55 | Number 1 Article 3

12-20-2018

Economic Rationality and Ethical Values in Design-Defect Analysis: The Trolley Problem and Autonomous Vehicles

W. Bradley Wendel

FINAL Wendell camera ready (Do Not Delete) 1/2/2019 10:27 AM

ECONOMIC RATIONALITY AND ETHICAL VALUES IN DESIGN-DEFECT ANALYSIS: THE TROLLEY PROBLEM AND AUTONOMOUS VEHICLES

W. BRADLEY WENDEL*

I. INTRODUCTION: MACHINE ETHICS AND THE TROLLEY PROBLEM

In the future with self-driving cars, the sensors and software of an autonomous vehicle may be confronted by a dilemma: crash into a telephone pole, killing the driver, or swerve into a crowd, killing five people. This may sound familiar if you have taken an introductory ethics class. Ethics teachers are fond of using the so-called trolley problem, introduced by Philippa Foot and further developed by Judith Jarvis Thomson,[1] to illustrate the difference between consequentialist and deontological moral theories. Moral theorists have used the trolley problem to explore subtle issues including the nature of intention, the doctrine of double effect (Foot's original use of the thought experiment), the act/omission distinction, agent-relative versus agent-neutral duties, moral agency and "authorship" of wrongs, and the problem of moral luck.[2] In the original version of the problem, the reader envisions herself in charge of a switch on a railroad track. A runaway trolley hurtles down the track, which, if left as is, will result in

* Professor of Law, Cornell University. The author gratefully acknowledges the research funding provided by the Judge Albert Conway Memorial Fund for Legal Research, established by the William C. and Joyce C. O'Neil Charitable Trust.

1. *See* PHILIPPA FOOT, *The Problem of Abortion and the Doctrine of Double Effect*, *in* VIRTUES AND VICES 19 (1978); Judith Jarvis Thomson, *Killing, Letting Die, and the Trolley Problem*, *in* RIGHTS, RESTITUTION, AND RISK: ESSAYS IN MORAL THEORY 78 (William Parent ed., 1986); Judith Jarvis Thomson, *The Trolley Problem*, *in* RIGHTS, RESTITUTION, AND RISK, *supra*, at 94.

2. *See, e.g.*, Michael Gorr, *Thomson and the Trolley Problem*, 59 PHIL. STUD. 91 (1990).

the trolley crashing into a minivan containing five people. By opening the switch, however, the reader can divert the trolley onto a section of track where a railroad employee is performing maintenance, resulting in his death. (As with many fanciful stories in philosophy, it is left unexplained precisely how this situation arose, and how the reader wound up in the role of trolley switch-attendant.) In any event, the issue is supposed to be clear: do nothing and passively allow the death of five people, or act and allow the death of one person while saving five others? If it seems clear that switching the trolley onto the track with one potential victim is the right thing to do, would the same principle apply to the situation facing a transplant surgeon who needs five organs to save five different patients, and who learns that in her hospital is a healthy person whose organs happen to be a perfect match for the five others otherwise certain to die?[3]

Many journalists and other commentators in the popular media have been captivated by the trolley problem and its myriad variations as they might arise for engineers designing autonomous vehicles. It has been described as "the focus of fierce debate among technologists around the world."[4] As an article in *Wired* puts it, "[p]eople seem more than a bit freaked out by the trolley problem right now."[5] *The Atlantic* reported on an ethical-engineering collaboration between Stanford University and driverless-car researchers at Google and Tesla using the trolley problem as a "useful springboard" for approaching the design of decision-making algorithms.[6] The problem does not arise only for fully autonomous ("driverless") vehicles,[7] but is implicated in semi-autonomous cars with present-generation technology, such as Tesla's

3. Thomson, *supra* note 1, at 80.

4. Alex Hern, *Self-Driving Cars Don't Care About Your Moral Dilemmas*, GUARDIAN (Aug. 22, 2016).

5. Aarian Marshall, *Lawyers, Not Ethicists, Will Solve the Robocar "Trolley Problem,"* WIRED (May 28, 2017).

6. *See* Lauren Cassani Davis, *Would You Pull the Trolley Switch? Does it Matter?*, ATLANTIC (Oct. 9, 2015).

7. The National Highway Traffic Safety Administration (NHTSA) distinguishes among levels of automation based on whether the driver or an automated system is primarily responsible for monitoring the driving environment. A vehicle may have multiple systems, some of which are highly automated, on NHTSA's definition, others of which are less highly automated. *See* U.S. DEPT. OF TRANSPORTATION, FEDERAL AUTOMATED VEHICLES POLICY 11 (Sept. 2016) [hereinafter "DOT AV Policy"].

Automatic Emergency Braking. Should such a system be programmed to take control away from a human driver if the car senses the driver is deliberately driving the car toward a group of people?[8] The topic has generated enough interest that a group of researchers sought to determine the preferences of survey respondents regarding an autonomous vehicle's solution to the trolley problem.[9] Should the car sacrifice one person to save the lives of five others? What if the one person in question is the driver? Perhaps unsurprisingly, study participants' preferences varied depending on whether they were asked to imagine themselves as the driver. Participants who assumed they would be in the driver's seat were considerably more likely to wish for a car that did not make the utility-maximizing calculation that one lost life was better than five. As a result, people may buy fewer self-driving cars, failing to mitigate the problem of crashes caused by driver error.[10] Perhaps aware of this finding, a Mercedes-Benz executive stated that if faced with the choice between running over a child who unexpectedly darted into the road and steering suddenly, causing a rollover accident that would kill the driver, an automated Mercedes would opt to kill the child.[11]

To be sure, not everyone thinking about the design of self-driving cars is obsessed with the trolley problem. A Google engineer working on that company's autonomous-vehicle program noted that if a vehicle found itself in the situation of having to choose between "the baby stroller or the grandmother" it would mean that there was a mistake a few second earlier in the accident sequence; thus, an ethical software engineer would concentrate on designing systems that minimize the likelihood of getting into a trolley-type dilemma in the first place.[12] In

8. *See* Patrick Lin, *Here's How Tesla Solves a Self-Driving Crash Dilemma*, FORBES (Apr. 5, 2017).

9. Jean-François Bonnefon, Azim Shariff & Iyad Rahwan, *The Social Dilemma of Autonomous Vehicles*, SCIENCE (June 24, 2016).

10. A *New York Times* article reports that 37,000 people died in car accidents last year in the U.S., "most from human error." David Leonhardt, *Driverless Cars Made Me Nervous. Then I Tried One*, N.Y. TIMES (Oct. 22, 2017); *see also* DOT AV Policy, *supra* note 7, at 5 (estimating that 94% of car accidents are the result of human error).

11. *See* David Z. Morris, *Mercedes-Benz's Self-Driving Cars Would Choose Passenger Lives Over Bystanders*, FORTUNE (Oct. 15, 2016).

12. Hern, *supra* note 4.

many of these scenarios, the right answer is always "slam on the brakes."[13] And, callous as it may sound, even if an automated system does occasionally make the wrong call, the net impact on safety is likely to be substantially positive given the improvement technology offers over error-prone human drivers.[14] The first pedestrian death involving

13. *Id.*

14. Leonhardt, *supra* note 10. In an otherwise engaging article, the author makes a statement that should not go unchallenged: "Technology creates an opportunity to save lives . . . . Just look at commercial airlines: Automation has helped all but eliminate fatal crashes among American air carriers. The last one happened in 2009." Technology has enabled aircraft manufacturers and airlines to engineer out risks that formerly contributed to accidents. Examples include Traffic Collision Avoidance Systems ("TCAS"), which has all but eliminated the risk of midair collisions in terminal-area airspace; Ground-Proximity Warning Systems ("GPWS"), which drastically cut down on controlled flight into terrain ("CFIT") accidents; and Predictive Windshear Alerting Systems ("PWS"), which reduced the risk of dangerous low-level windshear encounters. Better training for flight crews, particularly emphasis on Crew-Resource Management ("CRM") also played a role in reducing risk in commercial aviation. But it is a strongly-held belief in the aviation community that *automation*, as distinct from technology more generally, has created new risks of automation dependency that may have offset the reduction in risk attributable to automation. Inattention to or lack of proficiency in basic hand-flying skills is a contributing factor in several recent accidents, including the crash of Asiana Airlines flight #214 at San Francisco International Airport. *See* NATIONAL TRANSPORTATION SAFETY BOARD, ACCIDENT REPORT: DESCENT BELOW VISUAL GLIDEPATH AND IMPACT WITH SEAWALL ASIANA AIRLINES FLIGHT 214 (2013), https://dms.ntsb.gov/public/55000-55499/55433/563979.pdf. The problem of reversion to manual control is also an ever-present issue with automated control systems, as is the introduction of different types of programming errors. *See generally* Earl L. Wiener, *Cockpit Automation*, *in* HUMAN FACTORS IN AVIATION 433 (Earl L. Wiener & David C. Nagel eds., 1988).

This is a subject that is of central importance to the design of semi-autonomous cars and the liability of manufacturers. A fuller discussion must await a different occasion, but for now the point is that automation dependency is regarded as a serious risk in commercial aviation and needs to be taken seriously when designing interfaces between automated vehicles and human drivers. Curious readers are invited to Google "Children of the Magenta Line." American Airlines training captain Warren Vanderburgh coined that term to describe automation-dependent pilots who had forgotten that flight-management systems are tools for reducing workload at critical phases of flight, not substitutes for good old-fashioned piloting. The video in which he describes his experience flying the line with pilots who, quite understandably, had become overly dependent on automation is a bona fide classic in the aviation world. *See, e.g.*, Editorial, *How to End Automation Dependency*, AVIATION WEEK & SPACE

an autonomous vehicle, in March 2018, in a car conducting testing for Uber, did not appear to involve any decision by the car's software to sacrifice the pedestrian's life to save the driver or others.[15] Nevertheless, it appears that the trolley problem is now firmly established as a heuristic for an important normative issue connected with the design of autonomous and semi-autonomous vehicles. The issue is, what ethical standards should be employed when evaluating an engineering decision that plays out in an action taken by the vehicle, as opposed to the driver? Should autonomous systems make utility-maximizing decisions, or should they consider persons in general, or specifically the driver, as having inherent value? The trolley problem supposedly bears on this debate, at least to the extent it supposedly reveals the intuitions people have concerning the process of ethical decision-making.

II. AMORAL MACHINES?

In this paper I would like to make a modest, but pointed, intervention in the debate over the trolley problem and autonomous vehicles. A researcher at Stanford University recently declared the debate had already been concluded by, of all things, the law.[16] While moral philosophers were busy spinning out fantastic scenarios involving runaway trolleys and portly men being pushed off bridges (really, this is one variation in the literature),[17] lawyers and judges in

---

TECH. (July 19, 2013) (citing Capt. Vanderburgh's "famous lecture" on the paradox of automation).

15. *See, e.g.*, Sam Levin, *Video Released of Uber Self-Driving Crash That Killed Woman in Arizona*, GUARDIAN (Mar. 21, 2018); Heather Somerville et al., *Uber's Use of Fewer Safety Sensors Prompts Questions After Arizona Crash*, REUTERS (Mar. 27, 2018).

16. Bryan Casey, *Amoral Machines, or: How Roboticists Can Learn to Stop Worrying and Love the Law*, 111 NW. U. L. REV. 231 (2017).

17. *See* Thomson, *supra* note 1, at 82-83; DAVID EDMONDS, WOULD YOU KILL THE FAT MAN? THE TROLLEY PROBLEM AND WHAT YOUR ANSWER TELLS US ABOUT RIGHT AND WRONG 36-38 (2015). Although the stripped-down versions of trolley-type problems can seem silly, they are intended to track the features of very real moral dilemmas. Edmonds's popular history of the trolley problem begins with Winston Churchill's decision in 1944 to feed disinformation to the Germans which would foreseeably result in thousands of deaths in South London (from unguided German V-1 "buzz bombs") but save more lives, and protect government infrastructure, nearer to the center of the city. *Id.* at 1-7. President Harry Truman's decision to drop atomic

tort cases were busily establishing liability principles that answer the normative questions facing designers of autonomous vehicles. In this article, entitled *Amoral Machines*, the author contends that, while machine ethicists may fret over the balance of consequentialist and deontological considerations that an autonomous system should take into account when choosing between the driver and bystanders, autonomous-vehicle manufacturers will simply make decisions that minimize their exposure to legal liability.[18] They will act as the proverbial "bad man," described by O.W. Holmes, Jr., who cares about the law not because he believes it has any claim on his allegiance, but because he wishes to avoid legal sanctions.[19] Those decisions to prefer profit over ethics will result in algorithms that ignore negative externalities and take into account "only those costs the firm can expect to incur."[20] Depending on the applicable liability regime—strict liability, for example, as compared with negligence—designers will program vehicles differently, but always with the objective of minimizing the firm's liability. Ethical reasoning, however, is entirely beside the point.

This analysis in *Amoral Machines* is deeply confused in both matters of jurisprudence and tort law. The Holmesian bad man point of view is best understood as a caution not to interpret moralized language in the law (such as duty, reasonable care, good faith, and so on) as if it had the same meaning in the law as it does in ordinary life.[21] In this way, it is not a particularly profound insight, but merely a restatement of the central claim of legal positivism, that a norm is entitled to be called "law" because it was enacted by a particular social process, not because it is just or in furtherance of the common good. However, this does not mean that legal norms cannot track or incorporate moral principles. A lawyer would be well-advised to do some research to see what reasonable care or good faith requires under applicable legal

---

bombs on Hiroshima and Nagasaki, with the intention of ending the war early and avoiding the deaths of hundreds of thousands of combatants and civilians, has a similar structure. *Id.* at 23-24.

18. Casey, *supra* note 16, at 234.

19. *Id.* at 244 (discussing Oliver Wendell Holmes, Jr., *The Path of the Law*, 10 HARV. L. REV. 457, 459-62 (1897)).

20. *Id.* at 247.

21. *See* David Luban, *The Bad Man and the Good Lawyer*, 72 N.Y.U. L. REV. 1547, 1562-63 (1997).

standards, rather than simply consulting her moral compass. But it may turn out that what is *legally* required or permitted coincides with what is *morally* required or permitted. Tort law is full of examples of legal duties or privileges that make perfect sense from a moral point of view. The famous maxim "danger invites rescue,"[22] which is part of the law of proximate causation, is also a principle of folk morality: a tortfeasor who causes a primary injury is also liable to rescuers, because it is a normal human tendency to want to come to the aid of someone in peril. Implied assumption of risk cases may turn on an assessment of the social value of an activity, as in Judge Cardozo's entertaining opinion concerning the "Flopper" ride on Coney Island.[23] The analysis of legal duty is often explicitly moralized, as in the *Rowland*/*Tarasoff* factors employed by the California Supreme Court and picked up in many other jurisdictions.[24] The common law of punitive damages is replete with moral descriptions of the defendant's conduct as "wanton," "reckless," "consciously indifferent to the welfare of others," and is characterized by "malice," "spite," "ill will," or even a "disposition of perversity."[25]

22. Wagner v. Int'l Ry., 133 N.E. 437 (N.Y. 1921) (Cardozo, J.).

23. Murphy v. Steeplechase Amusement Co., 166 N.E. 173 (N.Y. 1929). *See generally* Kenneth W. Simons, *Murphy v. Steeplechase Amusement Co.: While the Timorous Stay at Home, the Adventurous Ride the Flopper*, *in* TORTS STORIES 207 (Robert L. Rabin & Stephen D. Sugarman eds., 2003).

24. *See* Tarasoff v. Regents of Univ. of Cal., 551 P.2d 334 (Cal. 1976); Rowland v. Christian, 443 P.2d 561 (Cal. 1968). For more recent applications, see Randi W. v. Muroc Joint Unified Sch. Dist., 929 P.2d 582 (Cal. 1997). For an application of the *Rowland*/*Tarasoff* approach by an important state appellate court outside of California, see J.S. v. R.T.H., 714 A.2d 924 (N.J. 1998). The California duty analysis requires a balancing of multiple considerations, including the foreseeability of harm to the plaintiff, the degree of certainty that the plaintiff suffered injury, the closeness of the connection between the defendant's conduct and the plaintiff's injury, the moral blameworthiness of the defendant's conduct, the policy of preventing future harm, the insurability of the liability, and the extent of the burden on the defendant and, indirectly, to the community as a whole. The California court's approach was at one time so prevalent in the analysis of the duty element that a leading torts treatise stated that a court's determination that a duty exists is nothing more than "an expression of the sum total of those considerations of policy which lead the law to say that the particular plaintiff is entitled to protection." W. PAGE KEETON ET AL., PROSSER AND KEETON ON THE LAW OF TORTS 357-58 (5th ed. 1984).

25. *See*, *e.g.*, Owens-Ill., Inc. v. Zenobia, 601 A.2d 633 (Md. 1992); National By-Products, Inc. v. Searcy House Moving Co., 731 S.W.2d 194 (Ark. 1987); Taylor v. Superior Court, 598 P.2d 854 (Cal. 1979). Unsurprisingly, Richard Posner proposed an economic rationale for punitive damages, wearing his hat as an appellate judge, in

The idea of a separation of law and morality is really a caricature of legal positivism, and not in any way an accurate description of the relationship between legal and moral considerations.

This is pretty bland stuff. A more radical reading of Holmes, picked up by some law and economics scholars, understands Holmes as contending that legal duties are not *duties* at all, no matter what the law says. The language of duty is instead a roundabout way of indicating that legislatures or courts have set a certain price on conduct as opposed to prohibiting it.[26] For example, the likelihood of being ordered to pay $250,000 in damages for ruining someone's reputation does not mean that slander is wrong and people should refrain from it; rather, it means only that it is an expensive activity. If one wishes to defame others and can afford to pay the price, the law has nothing more to say about the matter.[27] *Amoral Machines* appears to be relying on this reading of Holmes, but it is decidedly outside the mainstream among legal philosophers. For one thing, the reduction of legal prohibitions to prices or taxes cannot explain what would be wrong (if anything) with avoiding legal penalties by bribing judges, intimidating witnesses, or destroying evidence. The answer cannot be that *those* things are wrong, because the law-as-price view is supposed to apply to all purported legal duties. Moreover, as H.L.A. Hart pointed out, the radical reading of Holmes also fails to account for the fact that many people often treat the law as having normative significance.[28] That is, it establishes reasons for action, independent of any antecedent reasons one may have

---

an interesting case involving a hotel owner's indifference to the presence of bedbugs. *See* Mathias v. Accor Econ. Lodging, Inc., 347 F.3d 672 (7th Cir. 2003).

26. *See, e.g.*, Frank H. Easterbrook & Daniel R. Fischel, *Antitrust Suits by Targets of Tender Offers*, 80 MICH. L. REV. 1155, 1168 (1982) (arguing that rational corporate managers ought to treat legal norms as a form of price or tax, not an outright prohibition); Cynthia A. Williams, *Corporate Compliance with the Law in the Era of Efficiency*, 76 N.C. L. REV. 1265 (1998) (describing and criticizing the Easterbrook & Fischel position).

27. *See* Williams, *supra* note 26, at 1268. In fairness to Easterbrook and Fischel, they exclude *mala in se* criminal offenses such as murder and rape, so it is unclear what they would say slander. Maybe they think it is intrinsically wrong. In general, however, their version of the Holmesian bad man point of view denies that the law can actually prohibit anything. Conduct that is wrong in itself remains wrong; the law prohibiting the conduct does not change the normative situation of those subject to the law.

28. *See* H.L.A. HART, THE CONCEPT OF LAW 56, 82 (2d ed. 1994).

had to avoid punishment.[29] A theory of law that fits awkwardly at best with how the concept of *law* is used by ordinary folks is unlikely to be the right one.

Here is where the second, and in many ways more important, confusion arises. The author cites Learned Hand's famous B<PL formula from the *Carroll Towing* case, and alludes briefly to the position, developed by Guido Calabresi and Richard Posner, that the Hand formula embodies an economic analysis of the negligence standard.[30] The idea is that an actor compares the expected accident losses (PL) that could be prevented by the adoption of a burdensome safety feature, practice, or product redesign (B). If the cost of the additional safety precaution exceeds the savings in expected accident losses, the additional precaution is wasteful from the social point of view. Aggregate social welfare is maximized at the point at which the marginal cost of additional safety precautions is equal to the marginal benefit in terms of reduced accident costs. Never mind that juries are never instructed to engage in explicit cost-benefit calculation.[31] The more substantive problem with the economic reading of the Hand formula is that it fails to account for the results in a vast swath of torts cases. Courts simply do not impose liability only where the cost of a

29. Joseph Raz, *Authority, Law, and Morality*, *in* ETHICS IN THE PUBLIC DOMAIN 210 (1994).

30. Casey, *supra* note 16, at 242 n.66, 248. The classic Hand-formula case is *United States v. Carroll Towing Co.*, 159 F.2d 169 (2d Cir. 1947). Posner argued that Hand had given an economic formulation of the social function of the negligence standard, stated concisely: "Discounting (multiplying) the cost of an accident if it occurs by the probability of occurrence yields a measure of the economic benefit to be anticipated from incurring the costs necessary to prevent the accident." Richard A. Posner, *A Theory of Negligence*, 1 J. LEGAL STUD. 29, 32 (1972). The idea is that the law should require actors to prevent the accidents that are worth preventing—i.e. where the marginal cost of preventing them is less than the expected accident losses if a precaution is not taken. The total social cost of accidents, as the sum of prevention costs and accident costs, is thereby minimized. To put it differently, the level of spending on accident prevention is efficient if courts employ the *Carroll Towing* standard. *See* GUIDO CALABRESI, THE COST OF ACCIDENTS (1970); *see also* Guido Calabresi, *Some Thoughts on Risk Distribution and the Law of Torts*, 70 YALE L.J. 499 (1961).

31. *See* Patrick J. Kelley & Laurel A. Wendt, *What Judges Tell Juries About Negligence: A Review of Pattern Jury Instructions*, 77 CHI.-KENT L. REV. 587 (2002); Stephen G. Gilles, *The Invisible Hand Formula*, 80 VA. L. REV. 1015 (1994).

precaution is less than the accident cost savings.[32] A whole range of other factors comes into the analysis including an ethical assessment of the wrongfulness of the defendant's conduct; the defendant's duty being correlative to the *rights* of the injured party. Corrective justice and civil recourse theorists have for decades criticized economic analysis for failing to account for the centrality of the concept of moral wrongfulness in tort law.[33] A certain amount of rough, back-of-the-envelope-risk-utility analysis does go into evaluating when a defendant's conduct falls below the standard of care. Although juries may not be instructed on the Hand formula, reviewing courts will sometimes go through an impressionistic risk-utility balance,[34] though nothing like the kind of formal cost-benefit analysis required by administrative law. In addition to informal balancing, courts also take into account a wide range of factors that cannot be reduced to the cost of taking a precaution or the expected accident losses foreseeably prevented by the adoption of a precaution. Ordinary moral notions such as rights and wrongfulness are pervasive in tort law.[35]

Importantly for what will follow, economic analysis is inadequate to account for the role played by the principle of *responsibility*.[36] As

32. *See, e.g.*, Lawrence A. Cunningham, *Traditional Versus Economic Analysis: Evidence from Cardozo and Posner Torts Opinions*, 62 FLA. L. REV. 667 (2010).

33. *See, e.g.*, JULES COLEMAN, THE PRACTICE OF PRINCIPLE (2001); Gregory Keating, *Is the Role of Tort to Repair Wrongful Losses?*, *in* RIGHTS AND PRIVATE LAW 367 (Donal Nolan & Andrew Robertson eds., 2012); Mark Geistfeld, *Economics, Moral Philosophy, and the Positive Analysis of Tort Law*, *in* PHILOSOPHY AND THE LAW OF TORTS 250 (Gerald J. Postema ed., 2001); David G. Owen, *Philosophical Foundations of Fault in Tort Law*, *in* PHILOSOPHICAL FOUNDATIONS OF TORT LAW 201 (David G. Owen ed., 1995); John C.P. Goldberg & Benjamin C. Zipursky, *Torts as Wrongs*, 88 TEX. L. REV. 917 (2010); Ernest J. Weinrib, *The Special Morality of Tort Law*, 34 MCGILL L.J. 403 (1988).

34. *See, e.g.*, Washington v. La. Power & Light Co., 555 So.2d 1350, 1355 (La. 1990) (noting that the court's intuitive risk-benefit balancing "is merely a shorthand expression of the mental processes involved in such considerations" and that no court can "mathematically or mechanically quantify, multiply or weigh risks, losses and burdens of precautions.").

35. *See* Richard W. Wright, *The Standards of Care in Negligence Law*, *in* PHILOSOPHICAL FOUNDATIONS OF TORT LAW, *supra* note 33, at 249; Heidi Li Feldman, *Prudence, Benevolence, and Negligence: Virtue Ethics and Tort Law*, 74 CHI.-KENT L. REV. 1431 (2000).

36. COLEMAN, *supra* note 33, at 15.

Jules Coleman argues, economic analysis is a forward-looking theory of tort law[37] that establishes an efficient level of spending on accident prevention. There is no normative significance, in economic analysis, to the relationship between a wrongdoer and someone injured as a result of the wrongdoers acts or omissions. It would be a perfectly sensible thing to dispense with the category of tort liability altogether, and work instead with no-fault accident compensation (as in New Zealand), or a public-law idea such as social-risk regulation.[38] But we do not do this. American tort law is made up of many components parts, some substantive, others structural. There are doctrinal elements such as duty, standard of care, factual and proximate causation, and structural features such as the bilateral relationship between the plaintiff and the defendant, which is so obvious it often goes unnoticed. As Coleman points out, however, a significant feature of our tort system difficult to explain by an economic analysis of tort law is the fact that the victim sues the injurer, rather than the person who is in the best position to prevent accidents.[39] A number of misfortunes may befall a person in the course of her life. Only some of those misfortunes are legally actionable wrongs – that is, torts. Even on a fairly austere libertarian conception of responsibility, misfortunes deemed to be torts have the property, at least, of being attributable to human agency. "Volition and causation distinguish doings from mere happenings: actions from other events."[40] One may add further conditions, such as a requirement of moral or political fairness in the imposition of legal liability.[41] But a forward-looking explanation, reducing all concepts in tort law to accident-cost reduction, fails to make sense of the point and purpose of the tort system as it is revealed through its rules and application by judges and juries.

---

37. *See* Jules Coleman, *Tort Law and Tort Theory: Preliminary Reflections on Method*, *in* PHILOSOPHY AND THE LAW OF TORTS, *supra* note 33, at 183, 186.

38. Coleman, *supra* note 37, at 196.

39. *Id.* at 188; *see also* COLEMAN, *supra* note 33, at 17 ("There is simply no principled reason, on the economic analysis, to limit the defendant or plaintiff classes to injurers and their respective victims.").

40. Coleman, *supra* note 37, at 198; *see also* COLEMAN, *supra* note 33, at 51 ("Only agents are the proper objects of responsibility . . . .").

41. Coleman, *supra* note 37, at 200; *see also* George P. Fletcher, *Fairness and Utility in Tort Theory*, 85 HARV. L. REV. 537 (1972).

A fallback position might be to contend that, whatever is true of tort law generally, products liability law either is, or ought to be, treated differently. Perhaps products liability is a distinctive domain of tort law in which economic values and cost-benefit analysis should predominate. Products liability law is parasitic on underlying tort concepts such as reasonableness and causation. It is not separate from torts in any meaningful sense. One might respond that manufacturers are *strictly liable* for introducing defective products into the stream of commerce. Thus, what may be true of a negligence-based liability scheme, including related principles of proximate causation, comparative fault, and allocation of damages among jointly responsible tortfeasors, may not hold within the domain of strict liability. Applied to autonomous vehicles, principles of strict liability may support the *Amoral Machines* view that manufacturers will make design decisions based solely on economic considerations. This is a mistake, too, but one that is understandable in light of the evolution of products liability law. As the next section will explain, courts talked about strict liability in early and influential products liability decisions, but as the law evolved, the substantive liability rules coalesced around what is essentially a negligence standard. The modern analysis of design-defect cases, reflected in Section 2(b) of the Third Restatement, is mostly indistinguishable from negligence. That does not mean, however, that cases engage in the efficiency analysis suggested by Posner's reading of *Carroll Towing*. Design-defect cases are informed by a cluster of values related to the utility of the product to the consumer, in light of the performance of the product, its safety features, and the expectations consumers have with respect to the product. A modern court applying well-developed principles of design-defect analysis would engage in a process of ethical decision-making that is far from the Holmesian perspective described in *Amoral Machines*. To make that claim stick, the next section will briefly explain how the law of products liability evolved into what it is today.

III. FROM NEGLIGENCE TO STRICT LIABILITY AND BACK AGAIN[42]

Every first-year law student knows that strict liability for defective products coalesced in the early 1960's, led by influential decisions from the California Supreme Court. New York Court of Appeals Judge Benjamin Cardozo's *MacPherson* decision had long since liberated plaintiffs from the requirement of showing privity of contract in a lawsuit against the manufacturer of a defective product, but the underlying cause of action was still one for negligence.[43] Negligence worked tolerably well in some cases, but in others, the plaintiff lacked sufficient evidence to show the way in which the manufacturer's conduct fell below the standard of care. Traditional common-law evidentiary doctrines like *res ipsa loquitur* were helpful to some plaintiffs, but as California Supreme Court Justice Roger Traynor recognized in an influential concurring opinion, the policy considerations that supported relaxing the plaintiff's evidentiary burden would also support recognizing strict liability for product manufacturers.[44] Manufacturers should be responsible for injuries caused by product defects because they are in the best position to prevent the hazard, and because the resulting accident losses can be shifted and spread over the population of consumers. The plaintiff is typically in a worse position, *vis-à-vis* the manufacturer, to develop evidence through investigation and discovery to show the cause of the accident. This reasoning supports the application of *res ipsa*, but Justice Traynor recognized that a manufacturer is also unlikely to be able to introduce evidence sufficient to rebut the inference of negligence, and as a result "the negligence rule approaches the rule of strict liability."[45] Justice Traynor's view became the majority position in 1962,[46] around the time that Professor William Prosser was writing

42. *See* Sheila L. Birnbaum, *Unmasking the Test for Design Defect: From Negligence [to Warranty] to Strict Liability to Negligence*, 33 VAND. L. REV. 593 (1980).

43. *See* MacPherson v. Buick Motor Co., 111 N.E. 1050 (N.Y. 1916).

44. Escola v. Coca-Cola Bottling Co., 150 P.2d 436, 440 (Cal. 1944) (Traynor, J., concurring).

45. *Id.* at 441.

46. Greenman v. Yuba Power Prods., Inc., 377 P.2d 897 (Cal. 1962). We can set aside for present purposes the parallel developments in the law of contract warranties, which had also served as a doctrinal hook for plaintiffs to bring products

influential law review articles arguing for strict liability for manufacturers of defective products.[47] Professor Prosser was then serving as the Reporter to the American Law Institute's Second Restatement of Torts, and drafted the provision that became Section 402A. It provided that anyone who "sells a product in a defective condition unreasonably dangerous to the user or consumer" is strictly liable for the resulting harm.[48]

It did not take courts long to realize the problem lurking in this formulation. A manufacturer's liability is predicated on the finding of a *defect* in the product or, in the language of the Second Restatement, a conclusion that the product is in a defective condition unreasonably dangerous to the user. Not every dangerous product is defective. I own a heavy, sharp 10-inch cook's knife, which poses a risk to the fingertips of an unwary user, but it is not legally defective. In drafting Section 402A, Prosser anticipated this issue and included language in the commentary that distinguished dangerous products from those that are defective: "The article sold must be dangerous to an extent beyond that which would be contemplated by the ordinary consumer who purchases it, with the ordinary knowledge common to the community as to its characteristics."[49]

My cook's knife is just as dangerous as an ordinary consumer would expect, given common knowledge of its characteristics. The idea seemed to be that a defective product is one that—based on the ordinary consumer's knowledge and expectations—is *unreasonably* dangerous. However, even this interpretation seemed to be a step back from the strict liability envisioned by the Second Restatement. The California Supreme Court thought so, and cautioned that courts should not

---

liability claims against manufacturers. In a watershed case, the New Jersey Supreme Court recognized a cause of action for breach of the implied warranty of merchantability, but without the usual contract-law limitations of privity and disclaimability. The warranty attaches to the product upon sale and would be breached by a product that fails to perform as a reasonable consumer would expect. *See* Henningsen v. Bloomfield Motors, Inc., 161 A.2d 69 (N.J. 1960). The implied-warranty heritage of products liability law persists in the form of the consumer-expectation test for design defect, a subject which will be considered below.

47. *See* William L. Prosser, *The Fall of the Citadel (Strict Liability to the Consumer)*, 50 MINN. L. REV. 791 (1966); William L. Prosser, *The Assault Upon the Citadel (Strict Liability to the Consumer)*, 69 YALE L.J. 1099 (1960).

48. RESTATEMENT (SECOND) OF TORTS § 402A (AM. LAW INST. 1965).

49. *Id.* cmt. i.

"burden[] the injured plaintiff with proof of an element which rings of negligence."[50] Requiring the plaintiff to prove *both* that a product was defective *and* that the defect rendered the product unreasonably dangerous appears to "ring of negligence." But the court also recognized that manufacturers should not be treated as insurers of their products, i.e. subjected to absolute liability, which differs from strict liability.[51] There must be a limitation beyond which a court will not permit a jury to conclude that a manufacturer should be liable. The court believed the proximate cause element would do the necessary analytical work, but proximate causation often turns on reasonable foreseeability, and the word "reasonable" surely rings of negligence in this context as well.

It became clear in hindsight that the early "ancestor" cases like *MacPherson* and *Escola* had all dealt with a particular type of defect, known in modern law as a manufacturing defect.[52] A manufacturing defect exists when a particular product (a token of a type, to use philosophical language) deviates from the manufacturer's prototype in a way that causes it to fail and harm the plaintiff. Buick Motor Co., the manufacturer in *MacPherson*, had presumably intended to use wood of sufficient strength for the car's wheel spokes. Somehow, however, a weaker piece of wood was introduced into the manufacturing process, and as a result, the spoke failed, causing the car to crash. Similarly, some unknown glitch in the manufacturing process caused the Coke bottle in *Escola* to explode in the plaintiff's hand; the problem was either in the process of the glass bottle manufacturer, Owens-Illinois, or the manufacturer of the finished product, Coca-Cola. Manufacturing defect cases are easy to prove. The plaintiff must show only that the product departed from its intended design.[53] The manufacturer's prototype will likely be discoverable and, in any event, it is a reasonable inference that the manufacturer had not intended the weak wooden spoke or the flaw in the bottle that made it susceptible to exploding when handled roughly. Important for tracing the history of so-called

50. Cronin v. J.B.E. Olson Corp., 501 P.2d 1153, 1162 (Cal. 1972).

51. *Id.*

52. *See* RESTATEMENT (THIRD) OF TORTS: PRODUCTS LIABILITY § 2(a) (AM. LAW INST. 1998).

53. RESTATEMENT (THIRD) OF TORTS: PRODUCTS LIABILITY § 2(a) & cmt. c (AM. LAW INST. 1998).

strict liability, it is entirely appropriate to speak of manufacturing defect cases as involving true strict liability. Once the plaintiff establishes a departure from the manufacturer's specifications and a causal connection with her injury, the manufacturer is liable. The plaintiff need not show that the manufacturer failed to use ordinary care. Proof of a design defect never "rings of negligence."

The situation is entirely different with respect to design defect cases. The California Supreme Court observed that "[a] defect may emerge from the mind of the designer as well as from the hand of the workman."[54] That is certainly true. However, the difference is that the defect does not consist of a deviation from the manufacturer's specifications. Rather, it is *the specifications themselves* that are "defective."[55] But what does this mean? An obvious example would be something like a large, powerful industrial machine that is not guarded to prevent the user from inserting a body part into the machine while it is operating.[56] The challenge in a case like this is to come up with a way of evaluating the specifications as defective without either (1) recognizing absolute liability or (2) creating a test that "rings of negligence."[57] Finding liability regardless of the reasonableness of the engineering design choices made by the manufacturer risks going beyond strict liability into the imposition of absolute liability.[58] The manufacturer would then be liable as an insurer; all the plaintiff would need to show is a causal connection between her use of the product and an injury.[59] If the test is, instead, that it would be unreasonable to fail to equip the machine with a guard or interlock system, the standard

54. *Cronin*, 501 P.2d at 1162.

55. *See* RESTATEMENT (THIRD) OF TORTS: PRODUCTS LIABILITY § 2(b) & cmt. d (AM. LAW INST. 1998).

56. *See, e.g.*, Jaramillo v. Weyerhaeuser Co., 906 N.E.2d 387 (N.Y. 2009) (involving an "open architecture" box-folding machine lacking an interlock device that would shut the machine down if an open space is accessed by the user).

57. *See* John W. Wade, *On the Nature of Strict Tort Liability for Products*, 44 MISS. L.J. 825 (1973).

58. *See, e.g.*, Beshada v. Johns-Manville Prods. Corp., 447 A.2d 539 (N.J. 1982) (imposing liability for failure to warn of risks that, at the time, a reasonable manufacturer would not have known).

59. *See, e.g.*, Heaton v. Ford Motor Co., 435 P.2d 806 (Or. 1967) (distinguishing absolute and strict liability). *See also* Wade, *supra* note 57, at 828 (noting that under absolute liability, the manufacturer of a match would be liable for anything burned in a fire started by the match).

would hark back to the idea of negligence as the failure to use ordinary care. Courts occasionally tried to end-run that objection by claiming the design-defect standard did not ring of negligence because "negligence focuses on the conduct of the manufacturer while strict liability focuses upon the product."[60] That revision of the test in terms of the subject of analysis—product versus manufacturer—ran into the obvious objection that a finding of a product's design defect would necessarily entail the conclusion that the manufacturer's conduct fell below the standard of care; a reasonable manufacturer would not release a product into the market that embodied an unreasonable design.[61] Some courts continued to maintain the fiction that there is a difference between scrutinizing the manufacturer's conduct, which rings of negligence, and evaluating the product for whether it is reasonably safe or, instead, defective.[62] Eventually, however, courts accepted that the best resolution of this issue was to emphasize the word "unreasonably" in the Second Restatement formulation that liability follows a showing that a product is in a defective condition if it is unreasonably dangerous to the user.[63]

It cannot be overemphasized that there now is a consensus among courts and commentators that design-defect analysis does not follow a true strict liability approach.[64] The Second Restatement test contained the seeds of its own collapse. Liability requires showing a defect, and in design cases, that means showing that the product is unreasonably dangerous, not simply that an injury was causally connected to the product's design. The Third Restatement's approach is to require the plaintiff to show the existence of a reasonable alternative design that

60. Davis v. Globe Mach. Mfg. Co., 684 P.2d 692 (Wash. 1984); *see also* Prentis v. Yale Mfg. Co., 365 N.W.2d 176 (Mich. 1984) (observing that this reframing of the question "may have served to confuse, rather than enlighten, jurors").

61. *See*, *e.g.*, Lecy v. Bayliner Marine Corp., 973 P.2d 1110 (Wash. Ct. App. 1999).

62. *See*, *e.g.*, Anderson v. Owens-Corning Fiberglas Corp., 810 P.2d 549 (Cal. 1991) (involving a claim for failure to warn, but presenting the same conceptual issue).

63. *See*, *e.g.*, Boatland of Houston, Inc. v. Bailey, 609 S.W.2d 743 (Tex. 1980); Barker v. Lull Eng'g Co., 573 P.2d 443 (Cal. 1978).

64. *See* James A. Henderson, Jr. & Aaron D. Twerski, *Achieving Consensus on Defective Product Design*, 83 CORNELL L. REV. 867 (1998).

would have prevented the plaintiff's injury.[65] Significantly, the Third Restatement has entirely given up the position that manufacturers are subject to strict liability in design cases:

> Assessment of a product design in most instances requires a comparison between an alternative design and the product design that caused the injury, undertaken from the viewpoint of a reasonable person. That approach is also used in administering the traditional reasonableness standard in negligence. The policy reasons that support use of a reasonable-person perspective in connection with the general negligence standard also support its use in the products liability context.[66]

On the Third Restatement's risk-utility test, a jury should evaluate the plaintiff's proposed alternative design to determine whether it is a *reasonable* alternative design (a RAD, as it is often called). A product is defective in design if the manufacturer's failure to incorporate a RAD renders the product not reasonably safe.[67] The word "reasonable," which undoubtedly rings of negligence, is all over the black-letter test and commentary on design defect. The Third Restatement's test requires the trier of fact to balance the relative advantages and disadvantages of the product as designed, as compared with the plaintiff's proposed alternative. It can consider a variety of factors, including[68]:

- The magnitude and probability of the foreseeable risks of harm posed by the manufacturer's design, including the user's ability to avoid the risk by the use of reasonable care.

65. RESTATEMENT (THIRD) OF TORTS: PRODUCTS LIABILITY § 2(b) (AM. LAW INST. 1998).

66. *Id.* cmt. d (citations omitted).

67. *Id.* cmt. f.

68. *Id.* I have paraphrased and elaborated on the Comment f factors in the summary in text. In doing so I have referred back to their original source. *See* Wade, *supra* note 57, at 837-38. I am using bullet points and not numbers because the numbering of the Wade factors does not correspond to the presentation of the factors in Comment f. Nevertheless, I want to make it clear that the Restatement test is very close to that proposed in Wade's 1973 article. *See* Richard L. Cupp, Jr., *Defining the Boundaries of Alternative Design Under the Restatement (Third) of Torts: The Nature and Role of Substitute Products in Design Defect Analysis*, 63 TENN. L. REV. 329, 340-43 (1996) (discussing relationship between Wade factors and Third Restatement design-defect test).

- Consumer knowledge and expectations regarding the product, which may have their source in the manufacturer's marketing efforts, may simply be based on ordinary experience and knowledge, or may be influenced by the instructions and warnings accompanying the product.[69]
- The relative advantages and disadvantages of the product as designed and as it could have been designed, with consideration of the effect of the alternative design on product longevity, maintenance, repair, and aesthetics; for example, with due consideration for the hassle sometimes created by safety features and the natural tendency of users to disable those that create a huge headache.
- The range of consumer choice among products; for example, considering the appropriateness of a more dangerous but more versatile or useful version of the product intended for expert users.

The design-defect test can arguably be boiled down to one consideration stated in an influential law review article: "The manufacturer's ability to eliminate the unsafe character of the product without impairing its usefulness or making it too expensive to maintain its utility."[70]

On the modern design-defect test, my cook's knife is not defective in design because, although it poses significant foreseeable risks, there is no way to eliminate them without impairing the knife's usefulness. A cook's knife has to be heavy and sharp to perform its function of breaking down vegetables. A dull knife, obviously, would not be very useful. There are kitchen gadgets, such as mandolines and food

69. One has to be careful here because the open and obvious nature of the danger is *not* a sufficient reason for the manufacturer to avoid making a reasonable design change that would reduce the risk. *See* RESTATEMENT (THIRD) OF TORTS: PRODUCTS LIABILITY § 2 cmt. d, illus. 3 (AM. LAW INST. 1998). There may be cases, however, in which the consumer's knowledge that a product will perform in a particular way is a factor that goes into the analysis of whether the manufacturer has a duty to incorporate a safety feature. *See, e.g.*, McSwain v. Sunrise Med., Inc., 689 F. Supp. 2d 835, 842 (S.D. Miss. 2010) (holding that the wheelchair was not defective in design for failure to incorporate anti-tip tubes, because plaintiff and his father knew of the propensity of wheelchairs to tip backward and had chosen a model without anti-tip tubes).

70. Wade, *supra* note 57, at 837 (factor #4).

processors, that chop and slice vegetables and incorporate additional safety features, but they are not as versatile as a cook's knife, are a hassle to clean, and are often more expensive than a serviceable knife. Notice that the word "useful" here is not interpreted according to strict economic cost-benefit analysis. The issue is not the economic burden to the manufacturer in redesigning the knife. Rather, the analysis focuses on the user's experience with the product, including its "safety aspects . . . the likelihood that it will cause injury and the probable seriousness of the injury."[71] Balanced against the safety aspects of the product is not just the economic cost of a redesign but all of the disutility that would be associated with the design change, as considered from the user's point of view. The best product is not always the safest product once the design—as a whole—is evaluated in terms of its utility to the user.

Granted, the knife case would not involve a redesign but a wholesale replacement of the knife with a different sort of tool, such as a mandoline. However, the same analysis applies to a redesigned version of essentially the same product. Guards and interlocks on power tools can make them a hassle to use for certain applications. (I have often been tempted to wire shut the "deadman" switch on my lawn mower, which turns off the engine when the handle is released, because the mower must be restarted every time I have to leave it to pick up a stick or rock.) A canoe that is less likely to tip over may be slower and more difficult to paddle. Old-school automobile passive-restraint systems, such as automatic seat belts, drove many car owners spare until airbags supplanted the automatic belts.[72] Full-coverage body armor for law-enforcement officers is hot and restrictive, and may be less appealing to officers than vests that leave some gaps in coverage.[73] All of these examples illustrate the principle that the design-defect analysis considers the utility of the user experience, including expected accident costs, before and after the proposed redesign. The product is defective in design only if the plaintiff's proposed alternative is a RAD. This is known as the risk-utility test.

71. *See id.* (factor #2).

72. *See, e.g.*, Geier v. Am. Honda Motor Co., 529 U.S. 861 (2000) (reviewing the history of the Department of Transportation's Federal Motor Vehicle Safety Standard 208, which required manufacturers to equip vehicles with passive restraints starting in 1987).

73. *See* Linegar v. Armour of America, Inc., 909 F.2d 1150 (8th Cir. 1990).

It is necessary to conclude this historical overview with a brief note on a significant controversy in the development of the modern design-defect test. During the evolution of the design-defect standard, some states adhered to a freestanding consumer-expectation test, in which the expectations of consumers regarding product performance is not merely a factor to be considered in the risk-utility balance but is dispositive of the design-defect analysis.[74] In other words, a product was defective if it violated the reasonable expectations of consumers regarding its performance and safety. This test can be traced to the parallel development of implied warranties of merchantability in contract law, as an alternative doctrinal basis for holding the manufacturers of defective products liable to consumers.[75] It is also rooted in Section 402A of the Second Restatement in which the "defective condition unreasonably dangerous" standard was elaborated in the comments as involving "a risk of physical harm to an extent beyond that contemplated by the ordinary consumer . . . with the ordinary knowledge about the product's characteristics."[76]

The trouble with the freestanding consumer-expectation test was soon picked up by courts.[77] The issue was, not surprisingly, related to the word "reasonable." The consumer-expectation test cannot be based on the actual, subjective expectations of the plaintiff, but on those of a hypothetical reasonable consumer.[78] But then the question becomes, what expectations would a reasonable consumer have regarding product safety? A consumer could reasonably expect a pickup truck to be able to run over a one-inch rock without difficulty, but what about a six-inch

74. *See, e.g.*, Green v. Smith & Nephew AHP, Inc., 2001 WI 109, 245 Wis. 2d 772, 629 N.W.2d 727; Delaney v. Deere & Co., 999 P.2d 930 (Kan. 2000); Potter v. Chi. Pneumatic Tool Co., 694 A.2d 1319 (Conn. 1997). *See also* Aaron D. Twerski & James A. Henderson, Jr., *Manufacturer's Liability for Defective Product Designs: The Triumph of Risk-Utility*, 74 BROOK. L. REV. 1061 (2009) (providing a review of the case law and a defense of the risk-utility standard, by the Reporters to the Third Restatement).

75. *See* Henningsen v. Bloomfield Motors, Inc., 161 A.2d 69 (N.J. 1960). As a doctrinal basis for liability, implied warranty has been merged into the omnibus cause of action for defective products, which sounds primarily in tort. *See* RESTATEMENT (THIRD) OF TORTS: PRODUCTS LIABILITY § 2 cmt. n (AM. LAW INST. 1998).

76. RESTATEMENT (SECOND) OF TORTS § 402A cmt. i (AM. LAW INST. 1965).

77. *See* Henderson & Twerski, *supra* note 64, at 879-82.

78. *See, e.g.*, Campbell v. Gen. Motors Corp., 649 P.2d 224, 233 n.6 (Cal. 1982).

rock?[79] When it comes to complex product designs, inevitably involving tradeoffs among a number of functional and safety factors, a hypothetically reasonable consumer may have no expectations whatsoever. That consideration led the California Supreme Court, in an important case, to limit the consumer-expectation test to product designs in which ordinary experience is sufficient to permit the trier of fact to infer that the product failed to perform as a reasonable consumer would expect.[80] For example, a bus passenger who was injured in a fall was able to establish a design-defect claim using the consumer-expectation test because ordinary experience is sufficient to permit reasonable people to form expectations concerning the location and accessibility of handles and "grab bars" on a bus.[81] In complex design cases, however, neither the plaintiff nor the trier of fact has sufficient experience upon which to draw to evaluate whether the product violated reasonable expectations of safety and performance. As with any factual issue on which the trier of fact could be aided in understanding the evidence or determining a fact at issue, expert testimony may be introduced.[82] The expert's opinions will be based on the sorts of considerations taken into account by engineers or others who work on these product-design issues in the relevant industry. These considerations will generally be those captured in the Third Restatement's risk-utility analysis. In a case involving complex design issues, it is difficult to avoid the collapse of the consumer-expectation test into the risk-utility test.[83] On the Third Restatement's analysis, the

79. Heaton v. Ford Motor Co., 435 P.2d 806 (Or. 1967).

80. Soule v. Gen. Motors Corp., 882 P.2d 298 (Cal. 1994).

81. *Campbell*, 649 P.2d at 224.

82. FED. R. EVID. 702.

83. In a state purporting to employ the consumer-expectation test, the plaintiff's lawyer quite wisely opted to introduce expert testimony showing the feasibility of a design change to pneumatic power tools that allegedly would have prevented the neurological impairment to the plaintiff's hands. The plaintiff, in effect, proved the existence of a RAD using the risk-utility test as a kind of tacit background norm. *See* Potter v. Chi. Pneumatic Tool Co., 694 A.2d 1319 (Conn. 1997). The court paid considerable lip service to the consumer-expectation test, but the evidence in the case is better understood on a risk-utility framework. Subsequent case law in Connecticut, albeit unpublished and at the trial court level, supported this reading of *Potter*. *See* Gershberg v. Camera Wholesalers, Inc., No. FSTCV126014627S, 2014 WL 1283077 (Conn. Super. Ct. Feb. 26, 2014); Brierley v. Haas, No. WWMCV126005937S, 2014 WL 7714329 (Conn. Super. Ct. Dec. 18, 2014). Then the Connecticut Supreme Court

expectations of consumers regarding the product are a factor to be taken into account in design-defect litigation, but are not by themselves dispositive of the issue.[84]

IV. THE MORALITY OF DESIGN-DEFECT ANALYSIS

With this overview in modern products liability law, we can turn to the critique of the *Amoral Machines* thesis that autonomous vehicle manufacturers do not need to engage in ethical decision-making because the legal system has already resolved all of the relevant questions. The article sets up a classic trolley-type problem by imagining that the vehicle's software must instantaneously choose between running into five teenagers or engaging in a sudden evasive maneuver that will cause the car to overturn, killing its occupant.[85] On the "amoral" analysis, the resolution of the dilemma is that the car should avoid the teenagers, even if it results in the death of the driver. The first mistake in the analysis is to assume that the manufacturer will be strictly liable for the deaths of the teenagers, regardless of fault.[86] As explained in the previous Section, that is a complete mis-description of the governing liability standard. It confuses strict liability with absolute liability, which was never imposed.[87] Even in the early days

"modified" *Potter* (but really just clarified what had been implicit in the case all along, in my view) in a lengthy opinion. *See* Izzarelli v. R.J. Reynolds Tobacco Co., 136 A.3d 1232 (Conn. 2016).

84. For a good illustration of the role of consumer expectations in modern design-defect analysis, see Bourne v. Marty Gilman, Inc., 452 F.3d 632 (7th Cir. 2006). *See also* Vautour v. Body Masters Sports Indus., Inc., 784 A.2d 1178 (N.H. 2001) (explaining that the Second Restatement requirement of showing that a product is unreasonably dangerous to an extent beyond that which would be contemplated by an ordinary consumer is analyzed using a risk-utility balancing test).

85. Casey, *supra* note 16, at 242-43.

86. *Id.* at 243.

87. Going the other direction, the article also contends that if the teenagers were careless, they would be completely barred from recovery by the defense of contributory negligence. *See id.* at 243. Although a few U.S. jurisdictions continue to maintain contributory negligence as a complete defense, the overwhelming majority has moved to a rule in which the plaintiff's recovery is reduced in proportion to the share of responsibility allocated to the plaintiff by the trier of fact, but not barred completely. *See* RESTATEMENT (THIRD) OF TORTS: APPORTIONMENT OF LIABILITY § 7 & Reporters' Note (AM. LAW INST. 2000). Even in the few remaining true contributory-negligence (complete bar) jurisdictions there may be common-law

of products liability, when Section 402A of the Second Restatement was the applicable standard, a manufacturer would be held liable only for injuries caused by a *defective* vehicle.[88] Liability under Section 402A required a finding that the car was in a "defective condition unreasonably dangerous," which is not strict liability at all, but a kind of crypto-negligence standard. If the vehicle was not defective in design, then the manufacturer would not be liable, regardless of the causal connection between the design choices that went into the software and the deaths of the teenagers.

Fast-forwarding to the modern standard of the Third Restatement, Section 2(b), the question today would be whether the plaintiffs—here, the families of the teenagers killed by the car—can show that a reasonable alternative design (RAD) would have prevented the accident, and that the car was unreasonably dangerous due to the omission of the alternative design.[89] The plaintiff's proposed alternative design presumably would have been a different decision-making algorithm that would have swerved to avoid the teenagers, resulting in the death of the driver. Would it be a RAD? To answer that question, the trier of fact would be permitted to consider the factors set out above, developed from decades of caselaw and scholarly commentary. At this point, we can set aside the jurisprudential issue raised by the Holmesian bad man point of view. *Amoral Machines* contends that a profit-maximizing manufacturer will make product-design decisions based solely on the content of legal liability rules, not ethical principles:

> Robotics systems of the future will undoubtedly make decisions of immense ethical import. But their decisionmaking will be guided less

---

doctrines such as last clear chance that ameliorate the harsh effects of the rule. *See id.* § 3. Thus, the claim in the article that the causal contribution of the teenagers' behavior is sufficient to preclude their recovery is wildly inaccurate. In most jurisdictions the factfinder would be required to assign a percentage of responsibility for the accident to the teenagers, based on a comparison between their risk-creating conduct and that of the auto manufacturer. *See id.* § 8. In the minority of contributory negligence states, the factfinder would first have to determine whether the teenagers had been negligent (although this was stipulated in the hypothetical) and, if so, whether an exception such as last clear chance applies.

88. RESTATEMENT (SECOND) OF TORTS § 402A (AM. LAW INST. 1965).

89. RESTATEMENT (THIRD) OF TORTS: PRODUCTS LIABILITY § 2(b) (AM. LAW. INST. 1998).

> by the vagaries of "conscience" than by the "prophesy" of profit. These robots will view the world not as good moral philosophers, but as bad men – concerned less with idealized "ethical rule[s]" than with the legal rules that dictate whether their firms will face public sanction. And those that are instead engineered to follow "a clear and consistent" moral code will behave irrationally under a legal code lacking both such qualities.[90]

Whatever one thinks about the Holmesian bad man perspective generally, it should be clear that *if* the governing liability regime permits a court to impose liability on the ground that the manufacturer's design decisions do not embody the right ethical stance with respect to consumers, then a purely profit-maximizing manufacturer would seek to predict and respond to the ethical reasoning that a judge or jury is likely to employ. And it should also be clear that the factors set forth in Comment f to Section 2, the Wade article from which they are derived,[91] and the voluminous caselaw applying the risk-utility test that the applicable liability standard does not sidestep the ethical issue. In fact, it poses it squarely.

Amoral Machines further asserts that "[s]ystems optimized for profit will not fret over negative externalities, but only those costs the firm can expect to incur."[92] It is not much of an exaggeration to say that, on the law and economics account, the entire point of tort law is to force a profit-maximizing firm to internalize negative externalities as a way of incentivizing it to determine whether they could be avoided, at lesser cost, through the adoption of a safety precaution or product redesign.[93] Thus, a manufacturer had better fret over negative

90. Casey, *supra* note 16, at 244-45.

91. *See* Wade, *supra* note 57.

92. Casey, *supra* note 16, at 248. Note that this sentence makes a factual claim about what manufacturers will do, not what they ought to do. One long-standing criticism of economic analysis is that whatever methodological merit there is in making simplifying behavioral assumptions about human motivation, a further normative argument is necessary to support the conclusion that people ought to behave as self-interested utility maximizers. *See, e.g.*, Ronald Dworkin, *Is Wealth a Value?*, *in* A MATTER OF PRINCIPLE 237 (1985). For present purposes, however, I am willing to bracket this issue because it is quite clear that applicable legal standards require the manufacturer to engage in ethical reasoning, if only to predict and thereby minimize its exposure to legal liability.

93. *See, e.g.*, LOUIS KAPLOW & STEVEN SHAVELL, FAIRNESS VERSUS WELFARE 102-03 (2002) (explaining that the optimal liability rule is the one that minimizes the

externalities because the applicable liability rule may shift those costs to the manufacturer. The issue is, of course, when the design-defect test shifts the costs of injuries to the manufacturer. The analysis begins with our old friend, the reasonable person. The Third Restatement is perfectly clear that the evaluative standpoint from which design-defect claims are to be evaluated is the same as the perspective from which negligence claims are considered:

> Assessment of a product design in most instances requires a comparison between an alternative design and the product design that caused the injury, undertaken from the viewpoint of a reasonable person. That approach is also used in administering the traditional reasonableness standard in negligence. The policy reasons that support use of a reasonable-person perspective in connection with the general negligence standard also support its use in the products liability context.[94]

---

sum of accident costs and precaution costs); WILLIAM M. LANDES & RICHARD A. POSNER, THE ECONOMIC STRUCTURE OF TORT LAW 6-9 (1987) (summarizing the economic theory of torts as a mechanism for internalizing social costs).

94. *See* RESTATEMENT (THIRD) OF TORTS: PRODUCTS LIABILITY § 2 cmt. d (AM. LAW INST. 1998) (citation omitted). The "in most instances" proviso in this Comment refers to the quite rare category of cases in which the plaintiff can establish defective design without introducing evidence of a RAD. Those cases are mostly limited to so-called manifestly unreasonable designs, which present a low social utility and a high degree of danger. *See id.* cmt. e. The standard example is an exploding joke cigar, for which no redesign is available that would provide the same prank characteristics without the risk of causing burns to the consumer. *See id.*, Illus. 5. A few cases have applied the manifestly unreasonable design category to less obviously ridiculous products, such as above-ground swimming pools. *See, e.g.*, O'Brien v. Muskin Corp., 463 A.2d 298 (N.J. 1983). The majority approach, however, is illustrated by a case involving a backyard trampoline. *See* Parish v. Jumpking, Inc., 719 N.W.2d 540 (Iowa 2006). In that case, the plaintiff did not offer proof of an alternative design but instead relied on the "manifestly unreasonable" category. The court cited Comment d to the Third Restatement, Section 2, which states that many common and widely distributed products, such as firearms, alcoholic beverages and, contra the previously cited *O'Brien* case, above-ground swimming pools, are dangerous and cannot be redesigned. However, quoting one of the Reporters to the Third Restatement, it limited the category of manifestly unreasonable designs to those in which a trier of fact would conclude that the product was "so bad, so very out loud bad, so very antisocial, that it would tug against the very grain of the way [the factfinder] was raise." *See id.* at 544 (quoting James A. Henderson, Jr., *The Habush Amendment: Section 2(b) comment e*, 8 KAN. J.L. & PUB. POL'Y 86 (1998)).

Now we have come full circle, back to the claim that the negligence standard, formulated in terms of what a reasonable person would do, can be reduced without distorting the economic cost-benefit analysis. As noted above, a reduction of the reasonable person standard to a literal application of the *Carroll Towing* B<PL formula is an inaccurate description of the way judges instruct juries to apply the negligence standard.[95] It seems extremely unlikely that juries instructed only very generally, and understood as applying community moral values,[96] would consider only the comparison between expected accident cost savings and the expense to the defendant of the untaken precaution.

Rather than focusing on the reasonable person standard in negligence, which has been debated since time immemorial, we can look more specifically at design-defect cases arising relatively recently after courts worked through the puzzles created by the formulation of the Second Restatement's test in Section 402A. Imagine two rival designs for a system of sensors, software algorithms, and control linkages incorporated into a semi-autonomous vehicle. Each design would deal differently with the same foreseeable situation. The situation is a car rounding a blind curve and suddenly happening upon a stalled vehicle with five occupants. The car's speed is such that it cannot stop in time to avoid a rear-end collision with the stalled vehicle. However, it has the option of swerving to the right, onto a sidewalk, at the cost of running over a pedestrian. Having foreseen this situation, engineers working for the automobile manufacturer have considered two rival designs, called Utilitarian and Deontological, each of which is technologically feasible.[97] (These names are problematic, for reasons to be discussed below.) The decision principle built into Utilitarian algorithms is to choose the action that will increase the total happiness of all persons (or sentient beings, if one wishes to modify the problem to involve a dog on the sidewalk).[98] The Deontological design involves

95. *See* Kelley & Wendt, *supra* note 31; Gilles, *supra* note 31.

96. *See, e.g.*, Catharine Wells, *Tort Law as Corrective Justice: A Pragmatic Justification for Jury Adjudication*, 88 MICH. L. REV. 2348 (1990).

97. The feasibility of an alternative design is a separate issue from whether the alternative design embodies a superior net balance of safety and utility. *See, e.g.*, Flaminio v. Honda Motor Co., 733 F.2d 463, 468 (7th Cir. 1984).

98. *See, e.g.*, J.J.C. Smart, *An Outline of a System of Utilitarian Ethics*, *in* UTILITARIANISM: FOR AND AGAINST 3, 30 (J.J.C. Smart & Bernard Williams eds., 1973). For present purposes we can set aside the considerable disagreement within

a somewhat more complex decision-making algorithm. Essentially, it boils down to regarding every potentially affected person as having intrinsically important rights and refusing to permit the vehicle to interfere with someone's rights as a means of promoting the greater good.[99] As a result, the rights of a bystander may function as a side constraint on the algorithm[100] and, consequently, considering an option to intentionally swerve and kill the pedestrian may be ruled out.

Why do the five occupants of the stalled vehicle not have similar rights? Here is where the act/omission distinction or the notion of moral luck may be invoked; the presence of the stalled vehicle is taken as a given unless the causal sequence of events is altered by an active intervention of the car's automatic systems. To put it differently, negative duties (to refrain from killing) are more stringent than positive duties (to act in ways that prevent foreseeable harms).[101] The crash that kills five people may be preferred because it was not causally connected

consequentialist moral theory concerning issues such as what is to be maximized (pleasure, preference satisfaction, or something else); whether outcomes should be evaluated by averages or aggregates; the relevance of distributional considerations, such as whether one ought to make the *ex ante* worst off individual better off; whether actions, rules, or something else should be evaluated for rightness; and whether one's duty with respect to the good is to maximize it or merely aim at a certain value threshold. *See, e.g.*, David O. Brink, *Some Forms and Limits of Consequentialism*, *in* THE OXFORD HANDBOOK OF ETHICAL THEORY 380 (David Copp ed., 2006). We can also ignore the substantial practical problem of establishing an average value of the wrongful-death and survival claims asserted by the foreseeable victims. In the real world it makes a great deal of difference whether the defendant kills someone with substantial future earnings rather than a person working in a relatively low-earnings occupation. *See* KENNETH R. FEINBERG, WHAT IS LIFE WORTH?: THE INSIDE STORY OF THE 9/11 FUND AND ITS EFFORT TO COMPENSATE THE VICTIMS OF SEPTEMBER 11TH (2006). There is also the gruesome but inevitable question of whether the victim died instantly or suffered prolonged, conscious agony prior to death. Many jurisdictions allow compensation for noneconomic damages resulting from conscious, pre-death pain and suffering. *See* David W. Leebron, *Final Moments: Damages for Pain and Suffering Prior to Death*, 64 N.Y.U. L. REV. 256 (1989). These considerations would make a difference to the behavior of even the strictest Holmesian amoral profit-maximizer but would be very difficult to account for *ex ante* in the design of autonomous vehicles.

99. *See, e.g.*, Amartya Sen, *Rights and Agency*, *in* CONSEQUENTIALISM AND ITS CRITICS 187 (Samuel Scheffler ed., 1988).

100. *See* ROBERT NOZICK, ANARCHY, STATE, AND UTOPIA 28-30 (1974) (adopting Nozick's formulation of the function of rights).

101. FOOT, *supra* note 1; Thomson, *supra* note 1, at 81.

to an intentional act of the system—it was a mere happenstance from the point of view of the judgment the software had to make.

A manufacturer attempting to predict its legal liability as a result of adopting one of the rival designs would go through the analysis of Section 2(b) of the Third Restatement, which asks whether the expected accident losses could be reduced or eliminated by the adoption of a RAD. Note that the Holmesian bad man point of view, in which the manufacturer cares only about legal liability, does not necessarily favor one of the rival designs over the other. If Utilitarian is a RAD for Deontological, then a Holmesian "bad manufacturer" might incorporate as a default option the taking of the pedestrian's life. If, however, Deontological is a RAD for Utilitarian, because the applicable liability rule places central importance on the rights of potential victims, then the decision-making algorithm would reach the opposite result. It is also extremely important to notice at this point that the comparison between the rival designs is not straightforwardly consequentialist. Not only would this analysis beg the question against the Deontological alternative, but it is not faithful to the factors underlying the RAD analysis in Section 2. The reasonableness term of the RAD analysis is not simply whether one life, instead of five lives, would foreseeably be lost as a result of the adoption of one design. If the reasonableness analysis is best understood as a principle of corrective justice, it may be the case that the applicable liability rule requires manufacturers to design autonomous vehicle systems to avoid violating rights, even at the cost of allowing harms to befall others. The vehicle's systems do not kill the five victims—they are merely allowed to die. However, this seems like a conception of rights that a corrective justice theorist need not accept. As Judith Jarvis Thomson puts it, "there is no prima facie duty to refrain from interfering with existing states of affairs just because they are existing states of affairs."[102] Rights being correlative with duties, the five occupants of the stalled vehicle may have a right to demand that the autonomous vehicle's systems redirect it into the pedestrian. As Thomson argues in another article about the trolley problem, no violation of rights is involved by an agent's redirection of an existing threat so that it takes one life rather than five.[103] The vehicle's control software would be changing the path of a threat that

102. Thomson, *supra* note 1, at 84.

103. *Id.* at 107-09.

arose exogenously to the vehicle's systems; it would not be creating a new threat and directing it at the pedestrian.

V. DESIGN-DEFECT LITIGATION IN THE REAL WORLD

Whatever one thinks of the attempts by Thomson and others to respond to trolley-type thought experiments, two points bear emphasis. The first is that the interesting ethical questions related to the trolley depend on much more than the number of potential victims. The transplant-surgeon and fat man variations on the trolley problem demonstrate that while the first level of analysis of the problem is about rights versus utility, one cannot avoid dealing with deeper questions relating to the nature of rights, the relevance of intentions, and the relationship between the agent and the victim.[104] Why might it be morally permissible for a bystander to turn the trolley onto the side track, with the foreseeable result of the death of a single person, while it would not be permissible for a transplant surgeon to harvest the organs of a healthy person to save the lives of five others? This is a hard question if one thinks it is a fundamental principle of ethics that all persons are of equal importance.[105] Regarding only strictly impersonal, impartial values as a source of ethical reasons fails to account for the intuitive differences between the original trolley case and the transplant-surgeon variation. Philippa Foot's article that introduced the trolley problem to ethics relies on the traditional principle of double effect.[106] Thomas Nagel similarly relies on the agent's intention to explain the idea that one should not kill one person even to prevent a number of other foreseeable deaths.[107] On Nagel's account, deontological reasons tell us not to aim at evil as a means to a good end.[108] The moral quality of an agent's actions depends on the agent's intentions, not the foreseeable, but not desired, results. That

104. *See, e.g.*, THOMAS NAGEL, THE VIEW FROM NOWHERE (1986).

105. *See id.* at 171.

106. FOOT, *supra* note 1, at 19-21.

107. NAGEL, *supra* note 104, at 178.

108. Nagel follows Foot in thinking that the doctrine of double effect is the best solution to the trolley problem. *See id.* at 179-82.

means that one cannot do the right thing in the wrong way.[109] Put simply, the issue is not numbers but the content of mental states.

The second point follows from the first: it would be a significant challenge for software designers and other engineers to handle a trolley-type scenario in the right way—that is to say, reasonably. It seems fairly straightforward to implement the Utilitarian design. Systems must detect the number of people threatened by alternative courses of action and choose the action which produces the greatest savings in expected accident losses.[110] Suppose the family of the pedestrian selected for death by the autonomous systems sues the manufacturer of the vehicle claiming the vehicle with Utilitarian algorithms is defective in design. Assuming reasonable minds could differ, the trier of fact will be instructed to determine whether omission of the Deontological algorithm renders the vehicle not reasonably safe.[111] This is an open-ended inquiry including not only factors such as the foreseeable risks of harm, but also the expectations arising from product portrays and marketing, the utility of the product to the user, and the overall safety of the product.[112] The bottom line of the design-defect analysis is that the plaintiff must demonstrate that the harm—here, the death of the pedestrian—was *reasonably* preventable.[113]

Thinking from the perspective of a juror deciding the issue, there does seem to be something sinister about autonomous systems altering the path of a vehicle to aim directly at a pedestrian. Some jurors may have a Kantian intuition that the pedestrian is being used merely as a

109. *See* BARBARA HERMAN, *Integrity and Impartiality*, *in* THE PRACTICE OF MORAL JUDGMENT 23 (1993).

110. It should also be noted that an autonomous decision-making system would, by necessity, be insensitive to some particular features of a situation. For example, a computer would have no way of knowing if the pedestrian is related to the driver. The object of ethical analysis is therefore likely to be the rules by which the autonomous system determines what the vehicle should do. In the language of ethical theory, the analysis of the vehicle designer's choices would likely be a form of indirect consequentialism, in which the value of alternative courses of action is assessed in terms of "the values of the rules of motives under which the action can be subsumed." Brink, *supra* note 98, at 384.

111. RESTATEMENT (THIRD) OF TORTS: PRODUCTS LIABILITY § 2 cmt. f (AM. LAW INST. 1998).

112. *Id.*

113. *Id.*

means to the end of saving the lives of the five,[114] and thus, the pedestrian's sacrifice is unreasonable. (For Kant, what it means to be reasonable is to serve the objectively grounded end of the welfare of others, because it is a duty to do so.[115]) In response, the software designers could appeal to the doctrine of double effect, which regards the pedestrian's foreseeable death as a regrettable means to the good end of saving five lives. The decision-making software did not have an evil intention, even though it intentionally redirected the path of the car knowing it would kill the pedestrian. However, this response would fail to account for the intuition that an autonomous system should sacrifice the five occupants of the stalled vehicle, if necessary, to protect the driver of the car.[116] In the lingo of moral philosophy, the identity of the foreseen-but-not-intended victim, whose welfare will be sacrificed to save five others, gives rise to an agent-relative reason for the driver to prefer his own welfare, but not an agent-neutral reason which would apply to everyone impartially.[117] Should the driver's agent-relative considerations be taken into account by the systems' designers, or should they be impartial (agent-neutral) among the welfare of all potentially affected individuals? The question of

114. *See* CHRISTINE M. KORSGAARD, *Kant's Formula of Humanity*, *in* CREATING THE KINGDOM OF ENDS 106 (1996).

115. *Id.* at 107-08.

116. Jean-François Bonnefon, Azim Shariff & Iyad Rahwan, *The Social Dilemma of Autonomous Vehicles*, SCIENCE (June 24, 2016).

117. The idea is that agent-relative, or subjective reasons belong only to the agent. I have reasons to favor my own interest, and the interests of my family and friends, which are distinctively *my* reasons. Others have reasons not to harm my children—these are agent-neutral reasons, grounding negative duties—but I have agent-relative reasons others do not share to promote the welfare of my children. *See, e.g.*, Brink, *supra* note 98, at 383. This is a fairly standard distinction in modern moral philosophy, although its centrality for ethics has been questioned in an influential article resisting the claim that all ethical value must ultimately be agent-neutral. *See* CHRISTINE M. KORSGAARD, *The Reasons We Can Share: An Attack on the Distinction Between Agent-Relative and Agent-Neutral Values*, *in* CREATING THE KINGDOM OF ENDS 275 (1996). The decision of Mercedes-Benz, mentioned at the beginning of the article, to program an autonomous vehicle to kill a child who darted into the road as a means of preventing a rollover accident that would take the life of the vehicle's driver, is an example of agent-relative reasoning in action. *See* Morris, *supra* note 11. Agent-relative reasons are not absolute, however, and it may still be an open question whether the lives of two, three, or more bystanders could permissibly be sacrificed to save the driver's life.

reasonableness thus conceals an issue of deep interest to moral philosophers concerning the types of reasons from which ethical decision-making should proceed.

The defendant in a design-defect case is the manufacturer, which in most cases (likely *all* cases involving mass-marketed automobiles) is a corporation. The identity of the defendant creates a conceptual difficulty in the application of ethical concepts such as agent-relative reasons, the doctrine of double effect, and the prohibition on aiming at evil. One may wonder how a jury is supposed to assess the intentions of a corporation, which is capable of legal responsibility but is not a natural person. "How do [jurors] determine whether a group has intent or responsibility when those very terms are usually associated with individual, sentient human beings?"[118] One answer is that jurors do not assess the intention of a fictional person, but the mental state of specific human beings such as designers and engineers.[119] Presumably, a plaintiff's lawyer could depose the lead systems-integration or software engineer who worked on an autonomous vehicle design to determine that person's intentions regarding a trolley-type scenario. But even that inquiry would not reveal whether the victim should be regarded as a regrettable side-effect of a morally permissible intention, thereby exonerating the manufacturer from liability. Sophisticated critics of utilitarianism emphasize the importance of the agent's integrity, understood as a commitment to a distinctive set of projects and relationships.[120] It is not an exaggeration to say that the trolley problem was invented and developed to focus on the character of these deontological reasons. Jurors would have no choice but to ascribe them to a human engineer or to an autonomous system because otherwise there is no way of making sense of the idea of acting reasonably. Intention is inescapable in ethical reasoning.[121]

---

118. VALERIE P. HANS, BUSINESS ON TRIAL: THE CIVIL JURY AND CORPORATE RESPONSIBILITY 79 (2000).

119. *Id.* at 85 (noting that jurors tend to focus on specific individuals rather than the missing corporate "person").

120. *See, e.g.*, KORSGAARD, *supra* note 117, at 282; NAGEL, *supra* note 104, at 168; Bernard Williams, *Persons, Character, and Morality*, *in* MORAL LUCK 1 (1981).

121. An additional complication is suggested by research showing that people's intuitions concerning the intentions of others—which presumably would be directly implicated in a design-defect case in which the intentions of the manufacturer's employees would be at issue—are influenced by the moral goodness of the outcome.

That reasoning task would be required unless one fell back on the utilitarian strategy of relying on a simplistic tally of foreseeable victims associated with two or more options.[122] This is not only impoverished ethical reasoning,[123] it is also a gross misinterpretation of the Third Restatement design-defect standard. The ultimate issue in contention in a design-defect case is whether the product embodies a reasonable balance between safety, including the protection of third parties, and utility to the consumer. There is no algorithm for making this determination. As long as there is sufficient evidence to create a triable issue of fact, the jury must balance factors such as the feasibility of an alternative design, the likelihood and gravity of expected harm, and the disadvantages of the plaintiff's proposed alternative design.[124] The number of potential victims associated with different branches in the decision tree is a factor to be considered as an advantage or disadvantage of the plaintiff's proposed alternative design, but it is not

See Joshua Knobe, *The Concept of Intentional Action: A Case Study in the Uses of Folk Psychology*, *in* EXPERIMENTAL PHILOSOPHY 129 (Joshua Knobe & Shaun Nichols eds., 2008). For example, the judgment of whether an actor is said to have brought about a side effect intentionally is influenced by whether the side effect is good or bad. *Id.* at 133. In the Mercedes case, referenced *supra* note 11, a jury might determine that the car's action of killing a child who accidentally wandered into the road is wrong, and, thus, the manufacturer's design decision was unreasonable, even though the car acted to save the driver's life. Why? Because the child's death is a bad outcome. This is not conclusive of the ethical analysis. People may be mistaken, and their intuitive response may be subject to criticism. The extent to which experimental evidence should matter to moral philosophers is a currently a contested issue, particularly given the perennial question of how one is to derive an "ought" from an "is." *See* Joshua Knobe & Shaun Nichols, *An Experimental Philosophy Manifesto*, *in* EXPERIMENTAL PHILOSOPHY, *supra*, at 3, 10; Kwame Anthony Appiah, *Experimental Philosophy*, *in* EXPERIMENTAL ETHICS: TOWARD AN EMPIRICAL MORAL PHILOSOPHY 7, 15-16 (Christoph Luetge et al. eds., 2014). Outside of philosophy, however, whether they are fully committed Holmesian bad men or not, trial lawyers would presumably be quite interested to know how jurors actually apply the concept of intentional action.

122. *See* KORSGAARD, *supra* note 117, at 283 (noting the familiar strategy of dismissing deontological considerations and "castigat[ing] people who spend their time on worthless activities as irrational").

123. *See* Bernard Williams, *A Critique of Utilitarianism*, *in* UTILITARIANISM: FOR AND AGAINST, *supra* note 98, at 77, 99 (criticizing utilitarianism not for giving the wrong answer in a case like this, but for believing that it is *obvious* what the right answer should be).

124. *See, e.g.*, Cal. Civ. Jury Instr. § 1204.

dispositive. A jury might believe it is obvious that an autonomous vehicle's systems should steer it toward a single bystander to save five others but may also believe that it is wrong to program a vehicle to deliberately take the life of a bystander to avoid a greater number of deaths. Nothing in the law of liability for design defects precludes either of these conclusions.

VI. CONCLUSION

Far from putting an end to the ethical dilemmas potentially encountered by autonomous vehicles, the law requires manufacturers to engage in ethical reasoning, even if they do not want to. The design-defect standard calls upon the jury to consider whether the manufacturer's design choices were reasonable. Despite the best efforts of law and economic theorists, the reasonableness inquiry cannot be reduced to economic cost-benefit analysis along the lines of Richard Posner's interpretation of the *Carroll Towing* formula. Design-defect analysis does involve balancing expected harms and utilities, but these quantities cannot necessarily be measured in dollar terms. The functionality, usefulness, aesthetics, and headaches associated with safety features are all part of the risk-utility analysis, even for relatively simple products. When it comes to a complex semi- or fully autonomous vehicle with integrated sensor, control, and decision-making systems, a wide range of factors will inform the reasonableness analysis. One hopes that trolley-type situations will be extremely rare, and that automated systems will be able to intervene farther back in the accident sequence to avoid the necessity of choosing between one life and many. If it comes down to it, however, the decision procedure embodied in the vehicle's software may determine who lives and who dies. That decision will be evaluated by ordinary people, acting as judges and juries, using whatever resources bear on the question of reasonableness. This is not an amoral domain at all, but a richly moralized one. Even manufacturers who are concerned only with minimizing their exposure to legal liability should think through the ethical issues presented in these unusual cases.