

JOURNAL ON EMERGING TECHNOLOGIES

© 2023 William H. Widen

ARTICLES

AUTOMATED VEHICLES, MORAL HAZARDS & THE “AV PROBLEM”

WILLIAM H. WIDEN

INTRODUCTION 1
A. Hypothetical Safety Rating Scale 3
B. AV Industry Messaging & Regulation 5
I. THE SAFETY PROPOSITION..... 7
II. APPROACHES TO THE MORAL HAZARD 8
A. Enormous Financial Pressures at Time of Decision .. 10
B. High Tech Attitudes Towards Decision Making 12
III. UTILITARIAN JUSTIFICATIONS 13
A. Additional Utilitarian Justifications for Deployment 15
IV. TWO ATTITUDES TOWARDS THE SAFETY PROPOSITION 16
V. THE PIVOT TO “EDUCATION” 23
VI. CHANCE SET-UPS 24
VII. NEURAL NETWORKS, BUDDHIST TEMPLES AND OSTRICHES .. 26
VIII. IMPORTANCE OF ALTERNATE TESTING METHODS 29
IX. STRUCTURES TO PROMOTE RESPONSIBLE DECISION MAKING 30
X. COMPARING THE TROLLEY PROBLEM TO THE AV PROBLEM .. 33
A. The Difference Between a Trolley Case and a Trolley Problem 33
B. Application of a Trolley Case to the AV Problem 35
C. Bystanders Versus AV Companies 37
XI. THE NEXT OF KIN QUESTION..... 38
CONCLUSION..... 40
POSTSCRIPT 43

AUTOMATED VEHICLES, MORAL HAZARDS & THE “AV PROBLEM”

WILLIAM H. WIDEN*

INTRODUCTION

The automated vehicle (“AV”) industry faces the following ethical question: How do we know when AV technology is safe enough to deploy at scale?¹

I call the search for an answer to this question the “AV Problem.” This problem needs an answer now, more so than other ethical issues for AV design raised by the famous “Trolley Problem” in ethics² or the results of MIT’s experimental philosophy poll about “Moral Machines.”³ We face issues similar to the AV Problem now on a smaller scale with current testing of automated driving technology on our public highways⁴ where high-profile fatalities involving automation technology already have

* Professor, University of Miami School of Law, Coral Gables, Florida. The author is grateful for demonstrations of object and event detection and response (OEDR) technology at Georgia Institute of Technology’s computer science laboratory, and conversations with Philip Koopman, Peter Lederer, James Nickel, Philip Nickel, Deep Samal, and Marilyn Wolf. This essay is a gently updated version of a SSRN posting originally made in 2021.

¹ See, e.g., Patrick McGee, *Robotaxis: Have Google and Amazon Backed the Wrong Technology?*, FIN. TIMES (July 19, 2021), <https://www.ft.com/content/46ff4fe4-0ae6-4f68-902c-3fd14d294d72> (“Since Google launched its self-driving car project in 2009, the biggest challenge has been one of technology: can it be safe enough to deploy at scale?”). Deployment consists of the distribution and sale of series production AV models to the public. Some states have different regulatory standards for testing and deployment. See, e.g., CAL. VEH. CODE § 38750 (West 2017); 2021 Cal. Legis. Serv. Ch. 277 (S.B. 500) (WEST) (2021 portion of 2021-2022 Regular Session, updating § 38750 as of Sept. 23, 2021). See generally, PHILIP KOOPMAN, HOW SAFE IS SAFE ENOUGH?: MEASURING AND PREDICTING AUTONOMOUS VEHICLE SAFETY (2022).

² Judith Jarvis Thomson, *The Trolley Problem*, 94 YALE L.J. 1395 (1985). The name “Trolley Problem” comes from an ethical dilemma where one must make a constrained binary choice to pull or not pull a lever to direct a trolley onto a track with one worker and away from a track with five, when either choice is fatal to the persons on the track. It is based on scenarios originally presented by Philippa Foot in 1967. See *infra* text accompanying notes 129-143.

³ See E. Awad, S. Dsouza, R. Kim, et al., *The Moral Machine Experiment*, 563 NATURE 59 (2018).

⁴ Greg Bensinger, *Why Tesla’s ‘Beta Testing’ Puts the Public at Risk*, N.Y. TIMES (July 30, 2021), <https://www.nytimes.com/2021/07/30/opinion/self-driving-cars-tesla-elon-musk.html>. The issue is particularly serious as more states, such as Pennsylvania, pass laws allowing for driver-less testing of AVs. See William H. Widen, *Highly Automated Vehicles & Discrimination Against Low-Income Persons*, 24 N.C. J. L. & TECH. 115 (2022).

occurred.⁵ While stories about failures of vehicle automation technology get headlines,⁶ AV companies aim to deploy the more complex SAE Level 3, 4, and 5 automated driving system technology as soon as late 2023.⁷ Indeed, Philip Koopman and I have argued elsewhere that Tesla already has deployed SAE Level 4 motor vehicles in violation of law by selling its Full Self-Driving (FSD) “beta” software.⁸

This essay considers the AV Problem through the lens of two registration statements filed with the Securities and Exchange Commission (“SEC”): a November 5, 2021, filing of a registration statement on Form S-1 for Aurora Innovation, Inc. (“Aurora”), a company that hopes to be a leader in systems for AVs, and an August 27, 2021, filing of a registration statement on Form S-4 for Reinvent Technology Partners Y, the predecessor to Aurora.⁹ The Registration

⁵ See, e.g., Bryan Pietsch, *2 Killed in Driverless Tesla Car Crash, Officials Say*, N.Y. TIMES (updated Nov. 10, 2021), <https://www.nytimes.com/2021/04/18/business/tesla-fatal-crash-texas.html>; Letter from the National Transportation Safety Board (NTSB) to the U.S. Dept. of Trans., Responding to NHTSA DEC. 3, 2020 ANPRM (Feb. 1, 2021) (discussing lessons learned from NTSB crash investigations). The bad headlines keep coming following a decision by the California Public Utilities Commission to allow Cruise and Waymo to expand the scope of their pilot robo-taxi programs. See, e.g., Russ Mitchell, *San Francisco’s North Beach Streets Clogged as Long Line of Cruise Robotaxis Come to a Standstill*, L.A. TIMES (Aug. 12, 2023, 1:20 PM PT), <https://www.latimes.com/california/story/2023-08-12/cruise-robotaxis-come-to-a-standstill>.

⁶ See, e.g., Fred Lambert, *Tesla Full Self-Driving Beta Runs into a Pole in What Could be the First FSD Accident Caught on Video*, ELECTREK (Feb. 4, 2022, 8:49 PM PT).

⁷ Aurora Innovation, Inc. has disclosed its plans for deployment of AV trucks in 2023. See Aurora S-1, *infra* note 9, at 83. This essay focuses on “fully autonomous vehicles” with capability at Level 3, 4 or 5 as designated by the Society of Automotive Engineers (“SAE”). See SAE, TAXONOMY AND DEFINITIONS FOR TERMS RELATED TO DRIVING AUTOMATION SYSTEMS FOR ON-ROAD MOTOR VEHICLES, J3016_202104 (Apr. 30, 2021) [hereinafter J3016]. Mercedes-Benz has approval for the sale of Level 3 vehicles in California and Nevada. Dan Mihalascu, *Mercedes Drive Pilot Level 3 ADAS Approved for Use in California*, INSIDEEVS (June 9, 2023, 6:37 AM ET), <https://insideevs.com/news/671349/mercedes-drive-pilot-level-3-ad-as-approved-use-california/>.

⁸ See William H. Widen & Philip Koopman, *Do Tesla FSD Beta Releases Violate Public Road Testing Regulations?*, JURIST (Sept. 27, 2021), <https://www.jurist.org/commentary/2021/09/william-widen-philip-koopman-autonomous-vehicles/>. The case of Tesla gets more problematic with each news story, including the revelation that Tesla FSD Beta equipped vehicles have been programmed to run stop signs in “rolling stops” for which over 50,000 vehicles now have been recalled. See NHTSA, Part 573 Safety Recall Report, 22V-037, Tesla, Inc. (Jan. 27, 2022).

⁹ See, e.g., Aurora Innovation, Inc., Registration Statement (Form S-1) (Nov. 5, 2021) [hereinafter Aurora S-1]; Reinvent Technology Partners Y, Amendment No. 1 to Form S-4 Registration Statement (Form S-4/A) (Aug. 27, 2021) [hereinafter Reinvent S-4]; (collectively the “Registration Statements”). This essay does not address subsequent SEC filings by the company.

Statements reveal a potentially significant material omission: they fail to disclose Aurora’s *own* standard for deploying AVs at scale. Development of technology satisfying a more stringent safety standard takes longer than the development of a technology meeting a lesser standard, yet Aurora must deploy AVs quickly for financial success.¹⁰ For this reason, Aurora’s deployment standard is material and its omission is a potential violation of securities laws.¹¹

A. Hypothetical Safety Rating Scale

A simple hypothetical safety rating scale based on the number of miles driven without a fatality, expressed on a scale of 1 to 5, illustrates the deployment standard problem.¹² Assume the average unimpaired human driver would rate a 3 on this scale. Computer drivers¹³ rate a 2 for modest safety improvement over human drivers, and a 1 for significant improvement in safety. A 4 represents a modest decrease in safety from a human driver, and 5 a significant decrease in safety. An AV company in Aurora’s position has two realistic choices: *Choice 1*—either keep its deployment standard vague to preserve options in case of financial exigency or *Choice 2*—build trust by announcing that deployment will only occur after it can justify a safety case that its AV technology rates a 1. A middling choice of deployment at a 3 rating, or perhaps a 2 rating on this hypothetical scale, would conflict with public expectations that AV technology will achieve a significant safety improvement and not be merely value-neutral. It is not realistic that an AV company would announce a goal of a status quo level of safety.

¹⁰ This is seen most clearly in the financial projections contained in the Reinvent S-4 used in the de-SPAC transaction to take Aurora public. Projections appear in S-4 registration statements, but not S-1 registration statements. *See infra* note 46.

¹¹ Philip Koopman and I have argued elsewhere that the failure to disclose a deployment standard is a misstep taken by the AV industry which provides a reason to withhold trust. *See* William H. Widen and Philip Koopman, *Autonomous Vehicle Regulation & Trust*, 27:3 UCLA J.L. & TECH. 169 (2022). Failure to disclose a stringent deployment standard is the most problematic from a securities law liability standpoint.

¹² Philip Koopman and I have argued elsewhere that a simplistic national fatality statistic is inadequate as a sole metric for deployment decisions. *See* Philip Koopman & William H. Widen, *Breaking the Tyranny of Net Risk Metrics for Automated Vehicle Safety* (2023) (unpublished manuscript submitted to Safety-Critical Systems e-Journal for publication in 2024).

¹³ I use the term “computer driver” to refer to an AV with an automated driving system or ADS which has a capability at Level 3, 4 or 5 per SAE J3016. Capabilities below Level 3 are referred to as “advanced driver assistance systems” or ADAS and include such things as cruise control. *See* J3016, *supra* note 7, at 5.

Choice 1 preserves a “harm now, benefits later” utilitarian justification for deploying at a rating of 4 or 5.¹⁴ If Aurora wants that option, ethics requires public disclosure now so that an informed public debate might begin.¹⁵ The public might not readily accept such a justification.

Choice 2 builds public trust if Aurora’s management must defend a safety case¹⁶ to rate its AV technology a 1 to an independent committee of its board of directors before deployment. Adopting corporate governance structures to protect the integrity of deployment decisions, combined with a robust corporate ethics code, would strengthen a commitment to deploy only when evidence justifies a claim that an overall safety improvement immediately follows deployment. Disclosing deployment standards in SEC filings provides additional practical incentives against making safety a secondary concern in the face of financial exigency.

A moral hazard exists because the corporate form used to operate Aurora’s business (like other AV companies) shields investors and management from personal liability for the consequences of any mistaken decision by Aurora to deploy AV technology at scale before it is safe to do so.¹⁷

¹⁴ This raises the question of whether to discount later benefits in comparison with earlier harms. Discounted utilitarianism is controversial because it involves assigning a lower weight to future harms and benefits than present harms and benefits. See, e.g., Adam Jonsson & Mark Voorneveld, *The Limit of Discounted Utilitarianism*, 13 THEORETICAL ECON. 19 (2018). Moreover, the uncertainty of future conditions renders calculations problematic. See F. P. Ramsey, *A Mathematical Theory of Savings*, 38 ECON. J. 543, 549 (1928) (noting the possibility of future wars and earthquakes destroying accumulated savings). In the AV model, present harms might function like savings, which would be spent on future benefits—but only on the assumption that the future allows for spending the accumulation. The promised future AV benefits may never materialize. Ramsey believed discounting was ethically indefensible. *Id.* at 543.

¹⁵ Engineering professional ethics standards require transparency. See IEEE STANDARDS ASSOCIATION, IEEE 7000-2021, IEEE STANDARD MODEL PROCESS FOR ADDRESSING ETHICAL CONCERNS DURING SYSTEM DESIGN (Sept. 15, 2021) (approved June 16, 2021) [hereinafter IEEE 7000] (available via purchase or subscription, on file with the author).

¹⁶ A safety case is a structured logical argument, supported by evidence, that a system will be acceptably safe. It is not the same as the misleading “safety case framework” announced by Aurora. *Aurora’s Safety Case Framework*, AURORA, <https://safetycasework.aurora.tech/gsn> (last visited Aug. 26, 2023).

¹⁷ See generally William H. Widen, *Corporate Form and Substantive Consolidation*, 75 GEO. WASH. L. REV. 237 (2007).

B. AV Industry Messaging & Regulation

The essence of the AV industry’s messaging to date has been: just trust us, we are smart, we will do the right thing.¹⁸ When pressed, they reference vague content-free standards such as “sufficiently safe.”¹⁹ But the automotive industry already has a trust problem. Situations like the VW diesel emissions standards fraud,²⁰ EPA mileage inflation,²¹ Takata airbags,²² the GM ignition switch recalls,²³ Toyota’s criminal fine for lying to the National Highway Transportation Safety Administration²⁴ and the famous Ford Pinto gasoline tank design²⁵ make trusting the AV industry problematic because it is seen as part of the automotive sector. The public already withholds trust more generally from the automotive industry based on infamous past incidents such as these.²⁶

¹⁸ See, e.g., Aurora S-1, *supra* note 9, at 76, 84.

¹⁹ See, e.g., *Id.* at 7. An interview with Chris Urmson, Aurora’s CEO, illustrates the rhetoric of “sufficiently safe” used by industry participants. See Jerry Hirsch, *Autonomous Vehicle Pioneer Urmson Talks About Safety and Risks*, TRUCKS.COM (June 22, 2020), <https://www.trucks.com/2020/06/22/autonomous-vehicle-risks-urmson/>.

²⁰ Clifford Atiyeh, *Everything You Need to Know about the VW Diesel-Emissions Scandal*, CAR AND DRIVER (Dec. 4, 2019), <https://www.caranddriver.com/news/a15339250/everything-you-need-to-know-about-the-vw-diesel-emissions-scandal>.

²¹ Paul A. Eisenstein, *Mileage inflation sparks EPA scrutiny of carmaker claims*, NBC NEWS (Nov. 5, 2012), <https://www.nbcnews.com/business/business-news/mileage-inflation-sparks-epa-scrutiny-carmaker-claims-flna1C6856038>.

²² Clifford Atiyeh & Rusty Blackwell, *Massive Takata Airbag Recall: Everything You Need to Know, Including Full List of Affected Vehicles*, CAR AND DRIVER (Apr. 21, 2021), <https://www.caranddriver.com/news/a14499263/massive-takata-airbag-recall-everything-you-need-to-know-including-full-list-of-affected-vehicles>.

²³ Clifford Atiyeh, *GM, After Six-Year Battle, Settles Another Ignition-Switch Lawsuit for \$120 Million*, CAR AND DRIVER (Mar. 28, 2020), <https://www.caranddriver.com/news/a31965015/gm-settles-lawsuit-ignition-switch-car-values>.

²⁴ Office of the Attorney General, *Justice Department Announces Criminal Charge Against Toyota Motor Corporation and Deferred Prosecution Agreement with \$1.2 Billion Financial Penalty*, Press Release No. 14-286, U.S. DEPT. OF JUSTICE (updated Oct. 8, 2014), <https://www.justice.gov/opa/pr/justice-department-announces-criminal-charge-against-toyota-motor-corporation-and-deferred>.

²⁵ Ben Wojdyla, *The Top Automotive Engineering Failures: The Ford Pinto Fuel Tanks*, POPULAR MECHANICS (May 20, 2011), <https://www.popularmechanics.com/cars/a6700/top-automotive-engineering-failures-ford-pinto-fuel-tanks>.

²⁶ See, e.g., *Trust Lacking in Automotive Industry: How It’s Working to Improve Customer Engagement*, EKOMI BLOG, <https://blog.ekomi.com/2018/03/22/trust-lacking-in-automotive-industry-how-its-working-to-improve-customer-engagement/> (last visited Aug. 26, 2023) (discussing Takata and Toyota); Anton R. Valukas, *Report to Board of Directors of General Motors Company Regarding Ignition Switch Recalls*, JENNER & BLOCK (May 29, 2014).

Announcing a meaningful deployment standard and supporting that standard with deployment decision procedures builds trust more effectively than naked appeals for trust.²⁷ Lack of either articulated deployment standards or protective corporate governance structures makes a stronger case for regulation.²⁸ In this context, a mere litany of corporate culture virtues in SEC filings, press releases and websites is insufficient.²⁹

The AV industry resists regulation, arguing that regulations will become outdated before becoming operational,³⁰ slowing technological progress, while simultaneously arguing a utilitarian case for early deployment. Yet, as will be shown, the AV industry's recent *Best Practices* statement³¹ reveals no statistically significant metrics, standards, or data to back up the utilitarian claim that current AV technology reduces highway fatalities. To fill this void, the AV industry conducts a mere public relations campaign using a 501(c)(3) tax-exempt organization to convince the public of AV technology's potential benefits.³²

This essay continues by analyzing Aurora's Registration Statements' disclosure to isolate the AV problem, the details of the moral hazard, and the potential securities law violations. It compares the AV

²⁷ See Mark Alfano & Nicole Huits, *Trust and Distrust in Institutions and Governance*, THE ROUTLEDGE HANDBOOK OF TRUST AND PHILOSOPHY (Judith Simon ed. 2020). An industry strategy to self-proclaim one's own trustworthiness is unlikely to succeed. *Id.* (noting that "[d]irectly insisting on one's own good intentions when one is not already perceived as honest is thus not suitable for building and gaining trust").

²⁸ Though this essay refers to some aspects of regulation, providing a complete summary of regulatory developments for AV technology is beyond the scope of this project. For an overview, see BILL CANIS, CONG. RESEARCH SERV., R45985, ISSUES IN AUTONOMOUS VEHICLE TESTING AND DEPLOYMENT 8 (updated Apr. 23, 2021) (observing that "[p]roponents of autonomous vehicles note that lengthy revisions to current vehicle safety regulations could impede innovation, as the rules could be obsolete by the time they take effect.") [hereinafter ISSUES IN AV TESTING]; Bryant Walker Smith, *Automated Vehicles Are Probably Legal in the United States*, 1 TEX. A&M L. REV. 411 (2014).

²⁹ "Aurora's values guide our work and culture and support our ability to deliver our mission. They set the tone for the way we operate, they define who we are and how we do things, and they guide us when we face difficult situations." Aurora S-1, *supra* note 9, at 84. Those values are: Operate with integrity; Focus; No jerks; Be reasonable; Set outrageous goals; and Win together. *Id.*

³⁰ ISSUES IN AV TESTING, *supra* note 28. Historically the automotive industry has been subject to less stringent regulation than other industries developing and using safety critical systems such as aviation.

³¹ See *infra* note 76.

³² See *infra* text accompanying note 97. Individual automakers supplement this coordinated advertising campaign with their own efforts. See, e.g., GENERAL MOTORS, PATH TO AUTONOMOUS, <https://www.gm.com/commitments/path-to-autonomous.html> (last visited Aug. 26, 2023).

Problem to the “Trolley Problem”³³ which many use to discuss ethical problems in AV design. It concludes by outlining how AV companies might guard against the moral hazard of premature deployment and suggests a role for government regulation.

I. THE SAFETY PROPOSITION

The Registration Statements reveal that successful implementation of Aurora’s business plan *in the long term* depends on the truth of the following proposition: *a vehicle controlled by a computer driver is safer than a vehicle controlled by a human driver* (the “Safety Proposition”).³⁴

In a description of risk factors, the Registration Statements suggest several measures for public acceptance of AV technology, including *an expectation for better-than-a-human driving performance*.³⁵ The registrations statement disclosures identify this metric as a business challenge to overcome but not a prerequisite for deployment. The Registration Statements provide no details about how or when Aurora proposes to make this measurement and how it relates to deployment.

The Safety Proposition makes a statement about physical probability. This statement is either true or false depending on the state of AV technology. We might infer its truth to a reasonable level of certainty given appropriate data. The public, however, might *believe* the Safety Proposition even if it is not true because belief does not equal truth. The industry works to convince the public of the benefits of AV technology³⁶ hoping the public will infer the truth of the Safety

³³ See Thomson, *supra* note 2.

³⁴ See, e.g., Neal E. Boudette, *Tesla Says Autopilot Makes Its Cars Safer. Crash Victims Say It Kills*. N.Y. TIMES, <https://www.nytimes.com/2021/07/05/business/tesla-autopilot-lawsuits-safety.html?smid=url-share> (Sept. 1, 2021). The industry refers to a “computer driver” as an “automated driving system” or ADS. See J3016, *supra* note 7, at 6. An ADS can perform the entire dynamic driving task on a sustained basis in a specific operational design domain. It is completely self-driving insofar as it may operate in its operational design domain without constant human supervision. Vehicles properly identified at SAE Levels 3, 4 and 5 have this capability.

³⁵ Aurora S-1, *supra* note 9, at 7; Reinvent S-4, *supra* note 9, at 29. Some argue that an expectation test such as this does not work in tort litigation over complex products. See Emily Frascaroli, et al., *Let’s Be Reasonable: The Consumer Expectations Test is Simply Not Viable to Determine Design Defect for Complex Autonomous Vehicle Technology*, 2019 J. L. & MOBILITY 53 (2019) (discussing the future viability of the consumer expectations test in AV litigation).

³⁶ See *infra* Sec. V. THE PIVOT TO “EDUCATION”, text accompanying note 79 and following.

Proposition from these potential benefits, but potential benefits do not guarantee the truth of the Safety Proposition any more than belief does. Ethical concerns arise most clearly in the case of harm that might occur in a window period in which AV technology is deployed while the Safety Proposition is false.

II. APPROACHES TO THE MORAL HAZARD

Aurora (and, indeed, any AV company) might take four different stances towards the potential moral hazard associated with the decisions it faces.³⁷ If Aurora picks *Choice 1*, it can either tell the public that deployment might occur when the Safety Proposition is false, or it might remain silent. Disclosing its true stance toward safety in the case of *Choice 1* may create serious public relations problems. Aurora's failure to amend either the S-1 or the S-4, while otherwise remaining silent, is a good indication that Aurora is preserving the option to deploy when it either has no idea about the truth of the Safety Proposition or it has reason to believe it is false. This is true because *if* Aurora's principles for ethical development of technology allow it to deploy its AV technology at scale when the Safety Proposition is false or its status unknown, *then* its acceptance of this possibility will not conflict with the financial interests of its investors.

This is a hazard for the public, but not a moral hazard for Aurora's management in the classic sense because this attitude towards safety does not conflict with a fiduciary duty to stockholders.³⁸ The idea of limiting factors directors may consider in decision-making traces back to *Dodge v. Ford Motor Co.*³⁹ where the court held "[a] business corporation is organized and carried on primarily for the benefit of the stockholders. The powers of the directors are to be employed for that end."

In limited cases under the corporation laws of some states, directors might consider the public interest, and not only the shareholders' interests.⁴⁰ But as a general matter, if investors want to form a company that has a dual mission of positively impacting society and making a profit for shareholders, state corporation laws allow for the

³⁷ See text accompanying notes 12-17 (describing choices faced by Aurora).

³⁸ For a traditional description of corporate self-interest, see Frederick G. Kemplin, Jr., *The Public Interest in the Corporation*, 64 DICK. L. REV. 357 (1959).

³⁹ *Dodge v. Ford Motor Co.*, 170 N.W. 668, 684 (Mich. 1916).

⁴⁰ See Marsha E. Flora, *Redefining Pennsylvania Corporate Law: Eliminating Corporate Directors' Fiduciary Obligations*, 96 DICK. L. REV. 231 (1992) (allowing directors to consider the public interest when evaluating a hostile take-over bid).

formation of public benefit corporations (PBC) which are separate and distinct from non-profit corporations.⁴¹ Directors of a PBC make decisions that balance the stockholders’ financial interests, the best interests of other stakeholders affected by the corporation’s conduct and the public benefit identified in the PBC’s certificate of incorporation.

If Aurora picks *Choice 2* and adopts the moral principle that it will not deploy AV technology when the Safety Proposition is false (or when it has no idea about its truth or falsity), then it might make express disclosure of its stringent principle for deployment or it might remain silent. The option to remain silent with *Choice 2* makes no sense for two reasons.

First, it would be a missed opportunity to create the public trust which Aurora strives to achieve. Second, the failure to disclose an operating principle which might make its financial success more difficult clearly violates a number of its securities law disclosure obligations: (i) to the PIPE investors⁴² who committed in a private placement to provide funding in exchange for shares in the surviving company in the merger transaction (a violation of Rule 10b-5 of the Exchange Act), (ii) to the shareholders of Reinvent Technology Partners Y (“RTPY”), the SPAC vehicle,⁴³ who voted to approve the merger transaction (a violation of Section 14(a) and Rule 14a-9 of the Exchange Act), and (iii) Aurora’s existing shareholders who will receive shares in the surviving company as part of the merger transaction (a violation of Section 11 of the Securities Act of 1933).

Morgan Stanley, the sole book-runner for the initial public offering of RTPY shares, also may have liability based on its deferred \$34.2 million underwriting compensation for placement of the original RTPY S-1 securities because that fee is only paid if RTPY subsequently enters into an acquisition transaction.⁴⁴ One securities law liability theory would treat Morgan Stanley as a principal in the merger transaction (and not merely as aiding and abetting) based on its receipt

⁴¹ For example, Delaware enacted public benefit corporation legislation in 2013. As defined, a “public benefit corporation” is a for-profit corporation that is intended to produce a public benefit or public benefits and to operate in a responsible and sustainable manner. Section 365 of the Delaware General Corporation Law addresses the duties of the directors of a PBC.

⁴² A PIPE investor is an investor who makes a private investment in public equity.

⁴³ A SPAC is a special purpose acquisition company which is used in a transaction to take an operating company public without that operating company registering shares on an S-1 in an initial public offering or IPO. See John Coates, *SPACs, IPOs and Liability Risk under the Securities Laws*, Public Statement, Acting Director, Division of Corporation Finance, SEC (Apr. 8, 2021).

⁴⁴ Reinvent S-4, *supra* note 9, at 113.

of deferred underwriting compensation at consummation of the acquisition of Aurora.

The danger period for an Aurora management team committed to *Choice 2*, and for the public on either choice, arises shortly before dates Aurora set for deployment of AV technology—late 2023 for commercial trucking, and 2024 for ride hailing. Aurora describes its corporate culture as one that “sets outrageous goals”⁴⁵ essentially admitting a grave risk that it will not meet its deployment targets. This language suits morale building at a company picnic but is worrisome when advertised as a positive factor in SEC disclosure.

A. Enormous Financial Pressures at Time of Decision

When Aurora must decide whether to deploy or delay for more development and testing, its management likely will face enormous financial pressure to deploy. First, Aurora has a multi-billion-dollar market capitalization, though it is losing money, and will continue to lose money near term, with no positive EBITDA until projected free cash flow materializes in 2027.⁴⁶ Note with caution, however, this 2027 date assumes that Aurora’s projections are realistic. In fact, recent empirical evidence shows that projections in de-SPAC transactions are demonstrably less accurate than projections in other transactions.⁴⁷ The author’s discussions with market participants revealed that even well-modeled projections only have relevance for about 18 to 24 months.

Second, the prospect of lawsuits for disclosure violations, as I suggest is potentially the case with Aurora, provides an added incentive to deploy as scheduled regardless of the truth of the Safety Proposition to avoid the liability which will follow financial failure. It is impossible to have projected 2027 as the first year for positive free cash flow without also setting a prior deployment date, which Aurora announced for late 2023. And it is impossible to set a projected deployment date without having also decided upon a technical standard for deployment. This suggests that Aurora must already know its minimum safety standard for deployment and that this standard allows for a “harm now, benefits later”

⁴⁵ Aurora S-1, *supra* note 9, at 84; Reinvent S-4, *supra* note 9, at 244.

⁴⁶ Reinvent S-4, *supra* note 9, at 30, 132. The Aurora S-1 does not contain projections presumably because the safe harbor protections under the Private Securities Litigation Reform Act extend to companies completing a merger (as in a de-SPAC registered on a Form S-4) but not for shares registered in an IPO as with the Aurora S-1.

⁴⁷ See Elizabeth Blankespoor, et al., *A Hard Look at SPAC Projections* (Nov. 1, 2021) (noting that “while 42% of firms met their one-year forecast, this percentage declines to 21%, 10%, and 0% of their two-, three-, and four-year forecasts”), forthcoming in *MANAGEMENT SCIENCE*, available at SSRN: <https://ssrn.com/abstract=3961848>.

justification for deployment when the AV technology rates a 4 or 5 on our hypothetical safety rating scale. The alternative is that Aurora and its financial advisors did not take care in preparing projections.

Third, Aurora’s public offering was priced at \$10 per share (as are all shares in a de-SPAC), and its recent 52-week trading range has been between \$1.10 and 3.79 per share.⁴⁸ The share price might be expected to increase upon a successful deployment, and collapse if projections are not met. A reasonable assumption on the financial information presented is that, Aurora will fail if deployment is materially delayed.⁴⁹

Publicly adopting a clear deployment standard before a crisis reduces the risk of a later decision adverse to public welfare and thus builds trust.⁵⁰ The Registration Statements do not provide this clarity, and instead make bald statements such as Aurora will “operate with integrity,” “we do the right thing,” and Aurora wants to “earn trust with everything we do.”⁵¹ Apparently, disclosure that Aurora has “no jerks” on its payroll supports the comforting thought that we will all “win together.”⁵²

The Registration Statements’ disclosure even confirms the grounds for the moral hazard when they state: “[w]e expect each other to use good judgment and *always have the best interest of the company and our partners in mind.*”⁵³ This, of course, is the general corporate fiduciary duty. But that very duty validates a management decision to take a chance on successful early deployment rather than likely economic failure due to a delay because the calculus of decision does not include

⁴⁸ At market close on February 4, 2023.

⁴⁹ There are two other alternatives to financial failure. If Aurora can develop its AV technology sufficiently to demonstrate a “proof of concept,” then another industry buyer might acquire Aurora to obtain its technology, or the proof of concept might be enough to secure another round of financing. But, as a stand-alone company, Aurora likely fails if deployment is delayed in any material way. An effort to show proof of concept focuses on demonstrating functionality and not safety, assuming any accidents during testing can be paid for and subsequently advertised as “fixed.” The cost of a few lives may not provide an adequate deterrent given the monetary stakes.

⁵⁰ See ROBERT NOZICK, *How to do Things with Principles*, in *THE NATURE OF RATIONALITY*, 3, 10 (Princeton Univ. Press 1994) (noting that “[a]nnouncing principles is a way to incur (what economists term) reputation effects, making conditions explicit so that deviations are more easily subject to detection”). Aurora claims that its culture and values guide them “when we face difficult decisions.” Aurora S-1, *supra* note 9, at 84. This statement has little force, however, without expressly stating its values with respect to the deployment decision.

⁵¹ Aurora S-1, *supra* note 9, at 84; Reinvent S-4, *supra* note 9, at 244. This kind of self-affirmation is not a recommended strategy to build trust. See *Trust in Institutions and Governance*, *supra* note 27.

⁵² Aurora S-1, *supra* note 9, at 84; Reinvent S-4, *supra* note 9, at 244.

⁵³ Aurora S-1, *supra* note 9.

considering general social welfare in a morally acceptable way.⁵⁴ It merely considers the costs of settlement payments made to compensate for injuries and death.

If Aurora's management gambles on early deployment and wins, the company might survive and eventually prosper. If the gamble fails, the investors are no worse off because delayed deployment results in failure. In either case, their investment is lost. Management has nothing to lose by gambling on deployment when facing a financial crisis and has much to gain by taking a risk. The public has a great deal to lose. This presents a sound case for regulation, particularly in the absence of additional steps by the AV industry.

B. High Tech Attitudes Towards Decision Making

Prevailing attitudes in high-tech business culture might exacerbate the dangers of decision making against the backdrop of a financial crisis. Developing safe AV technology requires that software engineers solve a problem with life and death consequences, a prospect not present with development of other software products. Mark Zuckerberg of Facebook famously stated that “[t]he biggest risk is not taking any risk. In a rapidly changing world, the only strategy that is guaranteed to fail is not taking risks.”⁵⁵ A former leader of Google's self-driving-car unit stated: “If it is your job to advance technology, safety *cannot* be your No. 1 concern . . . If it is, you'll never do anything. It's always safer to leave the car in the driveway. You'll never learn from a real mistake.”⁵⁶ Sheryl Sandberg, also of Facebook, stated that “Done is better than perfect.”⁵⁷ Software engineers often believe they can “patch their way to perfection.” As explained to the author by the head of the department of computing at the University of Nebraska, the tendency of the software engineer is to deploy, see what happens, and then update software to address problems that surface. Software engineers often test

⁵⁴ Compare Dan Pontefract, *Should Companies Serve Only Their Shareholders Or Their Stakeholders More Broadly?*, FORBES (May 9, 2016),

<https://www.forbes.com/sites/danpontefract/2016/05/09/shareholders-or-stakeholders>, with Stephen Bainbridge, *A Duty to Shareholder Value*, N.Y. TIMES (Apr. 16, 2015), <https://www.nytimes.com/roomfordebate/2015/04/16/what-are-corporations-obligations-to-shareholders/a-duty-to-shareholder-value>.

⁵⁵ Nick Unsworth, *How to Take Risks That Win (Almost) Every Time*, ENTREPRENEUR (Nov. 30, 2017), <https://www.entrepreneur.com/article/304725>.

⁵⁶ Charles Duhigg, *Did Uber Steal Google's Intellectual Property?*, THE NEW YORKER, (Oct. 15, 2018), <https://www.newyorker.com/magazine/2018/10/22/did-uber-steal-googles-intellectual-property>. These ideas fit with the so-called “safety third” culture.

⁵⁷ SHERYL SANDBERG, *LEAN IN: WOMEN, WORK, AND THE WILL TO LEAD* 139 (2013) (Knopf).

products by encouraging customers to use a “beta” version of a product, a practice for which Tesla has been criticized.⁵⁸ Both Microsoft and Linux distributions develop software this way.⁵⁹ But a failure of a Facebook page or open-source browser simply differs from failures on a highway where patching to perfection creates risks to public safety.

III. UTILITARIAN JUSTIFICATIONS

When AV technology advances to the point that the Safety Proposition is true to a reasonable level of certainty, a simple utilitarian calculus provides a possible justification for deployment at scale. Before then, justifications for deployment rest on shakier ground. Compare five fatalities for a specified distance of miles traveled with human drivers to an alternate world with computer drivers which experiences only one fatality for the same distance traveled.⁶⁰ Neither the Aurora S-1 nor the Reinvent S-4 disclosure reveals anything that might make these near-term calculations easier.

If AV technology is deployed at scale in the near term, the status of the Safety Proposition will likely amount to informed conjecture at best, falling short of proof of the Safety Proposition to a reasonable certainty by use of a rigorous safety case. Yet, other disclosures in the Registration Statements suggests we already know that AV technology provides significant public benefits by reducing highway fatalities.⁶¹ If the truth of the Safety Proposition to a reasonable certainty is unknowable at the time of any deployment, a public debate becomes even more critical. Society ought to make decisions such as this rather than a small group of persons biased by a financial interest.

The example above in which switching to computer drivers reduces fatalities from 5 to 1 conforms to the observation of Daniel Kahneman, a Nobel Prize winner for work in behavioral economics, who opined about AV technology: “[b]eing a lot safer than people is not going to be enough. The factor by which they have to be more safe than humans

⁵⁸ Bensinger, *supra* note 4.

⁵⁹ See, e.g., *Educator Programs: Microsoft EDU Technology Adoption Program (TAP)*, MICROSOFT, <https://learn.microsoft.com/en-us/training/educator-center/programs/> (last visited Aug. 26, 2023); *Debian Testing*, DEBIAN, <https://wiki.debian.org/DebianTesting> (last visited Aug. 26, 2023).

⁶⁰ Complications arise with any metric because multiple facts impact safety in positive and negative ways. Examples include consumers purchasing more vehicles at SAE Level 2 with better ADAS systems, education reducing intoxicated driving and passage of state laws legalizing marijuana.

⁶¹ See *infra* Sec. VI. TWO ATTITUDES TOWARDS THE SAFETY PROPOSITION at text accompanying note 69.

is really very high.”⁶² The Registration Statements fail to warn that human psychology may demand a much higher threshold for successful deployment than a mere positive balance of utilities favoring AV technology.⁶³ The Registration Statements do not consider how satisfaction of a higher standard might pose additional challenges for Aurora.

A utilitarian moral test only requires a slight positive balance of utilities. If machine drivers and human drivers were equally safe, deployment would be morally neutral.

The truth value of the Safety Proposition has potential regulatory dimensions because a deployment decision at scale might require a regulatory approval conditioned upon demonstration of a safety case showing the truth of the Safety Proposition to a reasonable certainty. For example, the German Ethics Code states that the primary goal of AV technology ought to be the promotion of safety and an overall positive balance of benefits against burdens.⁶⁴ This appears to be the standard that the National Highway Transportation Safety Administration would apply if it produced substantive regulation.⁶⁵

Some make the case for rapid deployment of AV technology on the utilitarian grounds that earlier deployment will save more lives⁶⁶ without sufficient data to support this claim. This utilitarian argument conveniently complements the needs of AV companies to meet financial urgency by bringing products and services to market sooner rather than later. Arguments of this sort either expressly, or by implication, identify highway deaths as an emergency which justifies a relaxed approach to regulation.⁶⁷

⁶² Tim Adams, *Daniel Kahneman: ‘Clearly AI is Going to Win. How People Are Going to Adjust Is a Fascinating Problem’*, THE GUARDIAN (May 16, 2021), <https://www.theguardian.com/books/2021/may/16/daniel-kahneman-clearly-ai-is-going-to-win-how-people-are-going-to-adjust-is-a-fascinating-problem-thinking-fast-and-slow> (reporting observations of Daniel Kahneman).

⁶³ A technology which performs identically with a human driver has a greater likelihood of accidents with “bad optics” before crash statistics even out with scale. This might create a perception that AV technology is less safe than its actual long-term performance.

⁶⁴ C. Luetge, *The German Ethics Code for Automated and Connected Driving*, 30 PHILOS. TECH. 547, 558 (2017) (the “German Ethics Code”).

⁶⁵ NHTSA, Framework for Automated Driving System Safety, Proposed Rules, 85 Fed. Reg. 78058 (Dec. 3, 2020).

⁶⁶ See, e.g., J.F. Bonnefon, A. Shariff & I. Rahwan, *The Social Dilemma of Autonomous Vehicles*, 6293 SCIENCE 352, 1573-76 (June 24, 2016).

⁶⁷ See William H. Widen, *Autonomous Vehicles, COVID Vaccines & The Narrative of Exceptions*, JURIST (Sept. 1, 2021, 9:00 AM), <https://www.jurist.org/commentary/2021/09/william-widen-autonomous-vehicles-covid-vaccines/>.

A. Additional Utilitarian Justifications for Deployment

Two additional utilitarian arguments justify deployment of AV technology at scale if the Safety Proposition is false at the time of deployment.

Benefits and burdens may have a time dimension. More fatalities today might result in fewer fatalities in the long run, thus justifying early deployment based on a belief that earlier deployment would give needed experience to improve AV technology over time, bringing us closer to a future with zero traffic fatalities. There is no assurance this will happen.⁶⁸ Moreover, using people as mere means to an end in this way conflicts with other social norms.

Another early deployment justification suggests that merely meeting human driver performance (or perhaps an even lower standard) is morally acceptable if other benefits, such as lower freight costs and faster delivery times, outweigh any increased loss of life. One often-mentioned benefit is the ability of AV technology to solve a projected shortage of long-haul truckers. The analysis becomes complex when negative consequences are factored in, such as reducing the number of organs available for transplant, the impact on the financial health of the auto insurance industry and harm to the livelihood of lawyers who pursue and defend accident claims. In a world in which vehicles have no steering wheels, there might not be enough serviceable vehicles to deliver food to supermarkets, take patients to hospitals and perform other essential services if a product recall or cybersecurity threat took many vehicles off the roads simultaneously. Many people value self-reliance in times of crisis, yet widespread use of AV technology reduces self-reliance.

I understand that some in the AV industry have floated these alternate justifications for early deployment in private conversations, even though relying on truth of the Safety Proposition at the time of deployment seems the only viable public justification.

⁶⁸ Questions about the propriety of discounting future benefits in a utilitarian calculation make use of a “harm now, benefits later” justification problematic. See *supra* note 14.

IV. TWO ATTITUDES TOWARDS THE SAFETY PROPOSITION

As background, SEC filings like the Registration Statements use a certain vocabulary to signal the appropriate weight to give to various statements. For example, the Reinvent S-4 cautions that much of its disclosure consists of “forward-looking statements.” In SEC speak, a forward-looking statement is not a current or a historical fact, nor a guarantee of performance, but rather a qualified prediction.⁶⁹ Such a statement may, but need not, be signaled by use of words such as “believe,” “could,” “expect,” “may,” “should,” “strive,” “would” and similar expressions. As an example, consider a passage from the Reinvent S-4 suggesting increased safety as a benefit.

Improved safety. Each year 1.3 million people lose their lives in vehicle accidents globally. In the United States, truck transportation is the industry occupation with the highest number of fatalities. Human factors, such as fatigue, distraction, or recklessness, are estimated to contribute to up to 94% of crashes. *Autonomous cars and trucks can mitigate these factors through constant and consistent attention to the driving environment and advanced sensing and perception technology.*⁷⁰

All the statements of historical fact are backed up by references to different studies (as indicated by the footnote anchors 8 through 10⁷¹ in the quote), showing that the United States has a safety problem on its

⁶⁹ See 17 CFR § 230.175 - Liability for certain statements by issuers (defining “forward looking statements”); see also 17 CFR § 229.10 - (Item 10) General (describing the SEC position on company projections). “The Commission believes that management must have the option to present in Commission filings its good faith assessment of a registrant’s future performance. Management, however, must have a reasonable basis for such an assessment.” *Id.* at (1) Basis for projections.

⁷⁰ Reinvent S-4, *supra* note 9, at 233 (emphasis supplied) (internal footnotes omitted).

⁷¹ This 94% number is frequently misinterpreted to mean that human error causes 94% of all crashes. See, e.g., Don Kostelic, *The 94% Error: We Need to Understand the True Cause of Crashes*, STREETS BLOG USA (Oct. 14, 2020), <https://usa.streetsblog.org/2020/10/14/the-94-solution-we-need-to-understand-the-causes-of-crashes/>. The actual report does not say that human drivers cause 94% of crashes. See Santokh Singh, *Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey*, NAT’L HIGHWAY TRAFFIC SAFETY ADMIN., REPORT NO. DOT HS 812 115, at 1 (Feb. 2015), <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812115>. The SEC disclosure uses the softer term “contribute” rather than stronger word “cause” which appears in various myths promulgated by the AV industry. Philip Koopman, *Autonomous Vehicle Myths: The Dirty Dozen*, EE TIMES (Oct. 22, 2021), <https://www.eetimes.com/autonomous-vehicle-myths-the-dirty-dozen/>.

highways. The operative sentence in italics simply says *can*, not *will*, without support. In an SEC filing, this has the legal effect of saying that Aurora's product “*might*” do this. The Reinvent S-4 reveals the supposed benefits of AV technology as merely a hope or wish, not a promise.

Disclosure of the standard Aurora will apply to itself to deploy AV technology at scale would not require a forward-looking statement. Indeed, such a standard must have been set to provide a basis for the financial projections in the Reinvent S-4. True believers in AV technology might consider short term losses acceptable to realize long term gains because additional accident data from early deployment *might* help refine and improve AVs after the initial deployment at scale. Aurora does not state its position on the ethics of early deployment even though it must know the answer.

Consider disclosure at the start of the Reinvent S-4:

Aurora's service is not yet *commercialized*, RTPY has identified numerous challenges throughout its diligence in order for such service to be *commercialized*, and there is no guarantee that Aurora's service will be *commercialized*.⁷²

The Reinvent S-4 uses ‘commercialized’ repeatedly to signal business risks that Aurora faces. A failure to “commercialize” Aurora's AV technology at scale certainly will cause Aurora's business plan to fail. But, the Reinvent S-4 never highlights the importance of the Safety Proposition to successful commercialization. To succeed long term, Aurora's AV technology must perform at a safety level that exceeds the safety level achieved by human drivers, likely by a very wide margin.

The Reinvent S-4 does caution that the public expects safety performance superior to human drivers.

Solving self-driving is one of the most difficult engineering challenges of our generation. The industry can be characterized by a significant number of technical and commercial challenges, *including an expectation for better-than-a-human driving performance, . . . a need to build public trust and brand image, and real world operation of an entirely new technology*.⁷³

⁷² Reinvent S-4, *supra* note 9, at xv (emphasis supplied).

⁷³ Reinvent S-4, *supra* note 9, at 29 (emphasis supplied); *see also* Aurora S-1, *supra* note 9, at 7.

The Reinvent S-4 mentions “better-than-a-human” driving performance only once but without identifying it as Aurora’s own standard for the deployment decision, merely suggesting that this expectation characterizes a general perception for success of the industry.

Many other statements link commercial success to a generic achievement of safety without specifying a standard or metric.

The successful development of our self-driving systems and related technology involves many challenges and uncertainties, including:

. . . *achieving sufficiently safe self-driving system performance as determined by us, government & regulatory agencies, our partners, customers, and the general public*; . . . finalizing self-driving system design, specification, and vehicle integration; . . . *successfully completing system testing, validation, and safety approvals . . .*⁷⁴

The Reinvent S-4 never expressly describes what standard Aurora’s AV technology must satisfy in Aurora’s view to achieve a “sufficiently safe self-driving system performance” other than indicating the standard will be determined by several different parties, including Aurora itself.⁷⁵ The Reinvent S-4 discloses a lack of uniform standards for AV performance but never tells you that the industry’s own *Best Practices* do not clarify how to determine whether a standard of better-than-a-human-driver has been met.

In March 2021, the Automated Vehicle Safety Consortium published the *AVSC Best Practice for Metrics and Methods for Assessing Safety Performance of Automated Driving Systems (ADS)*.⁷⁶ These metrics apply to commercial fleets of vehicles such as Aurora’s planned initial roll-out to the trucking business. Section 4 of this publication states:

⁷⁴ Reinvent S-4 at 29 (emphasis supplied).

⁷⁵ *Id.*

⁷⁶ AUTOMATED VEHICLE SAFETY CONSORTIUM, A PROGRAM OF SAE ITC, AUTOMATED VEHICLE SAFETY CONSORTIUM™ BEST PRACTICE, AVSC00006202103 (Issued 2021-03) [hereinafter BEST PRACTICES].

ADS developers and manufacturers should build evidence, including metrics, *to support the argument* that their ADS is acceptably safe to operate on public roads.⁷⁷

Notice that the metrics have been developed with a view towards *advocacy*—“to support the argument.” A worry about advocacy is a worry about controlling beliefs—constructing a narrative. Focusing on “arguments” differs from searching for the truth of the Safety Proposition. A footnote makes this advocacy role clear:

Above all, the safety case exists to communicate an argument. It is used to demonstrate how someone can reasonably conclude that a system is acceptably safe from the evidence available. Absolute safety is an unobtainable goal. Safety cases are there to convince someone that the system is safe enough (when compared against some definition or notion of tolerable risk).⁷⁸

References to “reasonably conclude,” “safe enough” and “convince” are worrisome.⁷⁹ In civil litigation a jury might arrive at a reasonable conclusion if it is supported by some evidence. A judge will not disturb a jury’s decision unless no evidence supports a verdict. That is the concern about a “reasonable” conclusion. We are never told what is “safe enough” other than it “should be compared against some definition or notion of tolerable risk” which is never identified.⁸⁰ This amounts to a content-free generic description of evaluation of a safety case. That standard ought to be the truth of the Safety Proposition—at least to “best judgment” and not merely a judgment with some support.

The *Best Practices* states two safety goals identified as “high level” for “desired societal impact”:

⁷⁷ *Id.* at 8 (emphasis supplied).

⁷⁸ *Id.* at 8 n.3.

⁷⁹ Note that “justify” differs from “convince.” A typical definition in keeping with current practice is: “a Safety Case is a structured argument, supported by evidence, intended to justify that a system is acceptably safe for a specific application in a specific operating environment.” The lesser standard of “convince” used in the *Best Practices* raises concerns.

⁸⁰ The BEST PRACTICES should pick one of the established frameworks for risk and indicate which calibration to use. See Phil Koopman, *Different Types of Risk Analysis: ALARP, GAMAB, MEMS and More*, BETTER EMBEDDED SYSTEM SW, (May 8, 2021), <https://betterembsw.blogspot.com/2018/09/different-types-of-risk-analysis-alarp.html>.

1. Reduce the number and severity of crashes.
2. Perform contextually safe vehicle motion control.⁸¹

The metrics continue: “at a global level, a positive risk balance is achieved when the combination of frequency and severity are reduced.”⁸² Nothing explains how to balance a decrease in frequency, with an increase in severity, or vice versa.

Presumably severity is related not only to the seriousness of the injuries, but also the number of persons injured or killed--and yet the *Best Practices* do not indicate an appropriate weighting or otherwise elaborate on determination of a “positive risk balance.” Indeed, a reduction in minor-to-moderate accidents might generate statistics suggesting overall safety improvement when the number of fatalities has increased. The *Best Practices* merely states “[s]ocietal harm is the combination of severity and frequency of the total crash population for the ODD of interest where the goal is to reduce both.”⁸³

Best Practices notes that the United Nations ECE WP 29 forum has chosen what the AVSC dubs a “safety argument” for Level 3 vehicles—that the technology is “free of unreasonable (foreseeable and preventable) safety risks.”⁸⁴ The problem with some aspects of AV technology, particularly software architecture using neural networks, is that unforeseen risks are certain but the details unpredictable.⁸⁵ An engineering approach to address the problem of this “known unknown” requires limiting the scope of the harm potentially caused by a generic unforeseeable risk in a neural network.

The *Best Practices* hope to satisfy something akin to the Safety Proposition. But there is a significant chicken and egg problem because “safety assessment methods and metrics are still evolving” and “need future revision or expansion based on knowledge and experience gained

⁸¹ BEST PRACTICES, *supra* note 76, at 9.

⁸² *Id.*

⁸³ *Id.* at 10.

⁸⁴ *See, e.g.*, Inland Transport Committee, World Forum for Harmonization of Vehicle Regulations, Framework document on automated/autonomous vehicles, UNECE (updated Nov. 2021). “The level of safety to be ensured by automated/autonomous vehicles implies that “an automated/autonomous vehicle shall not cause any non-tolerable risk”, meaning that automated/autonomous vehicle systems, under their automated mode (Operational Design Domain (ODD) or Operational Domain (OD)), shall not cause any traffic accidents resulting in injury or death that are reasonably foreseeable and preventable.” *Id.* at 2.

⁸⁵ *See, e.g.*, Zoe Porter, et al., *Unravelling Responsibility for AI*, arXiv:2308.02608v1 [cs.AI] (Cornell Univ. Aug. 3 2023) (noting that the intended functionality of machine learning systems cannot be explicitly specified, with inherent complexity, uncertainty and opacity), <https://doi.org/10.48550/arXiv.2308.02608>.

following wider-spread deployment of [AV] technology.”⁸⁶ No one may know to a reasonable certainty whether Aurora’s AV technology has met this standard prior to a near term deployment. An evaluation of whether a computer driver is safer than a human driver must await data from more actual highway performance.

A footnote confirms that “[a]chieving statistically significant results for safety-relevant events across the entire ODD [Operational Design Domain] can require a significant time delay (i.e. for safety outcomes),” though accidents can provide “timely insights into safety performance.”⁸⁷ The idea is to measure the “safety impact of [AV] deployment over time to quantify progress toward meeting safety goals.”⁸⁸ Without a sufficient number of real accident cases involving AVs to evaluate, it is hard to build a safety case with statistical relevance, particularly in the short term.

To compound the problem, the conclusions of any successful presentation of a safety case do not apply when a vehicle operates outside its Operational Design Domain (ODD). Operation of an AV must be confined to an ODD to have confidence that the conclusions of a safety case presentation are relevant (even with sufficient data to make a safety case in a limited environment). One may not easily defend a metric when an accident occurs outside an ODD if that accident situation is one anticipated by publications such as ANSI/UL 4600, Standard for Safety for the Evaluation of Autonomous Products.⁸⁹

The Reinvent S-4 describes technology developments which contribute to safety:

We believe these early investments in our technology will enable us *to move toward commercialization more safely and quickly* than would otherwise be possible. . . .

The successful execution of these details of self-driving technology is what we believe will allow us to differentiate ourselves by developing *leading self-driving technology that can safely and reliably navigate its environment*.⁹⁰

⁸⁶ BEST PRACTICES at 4. In other words, one needs to deploy in order to find out whether deployment was safe in the sense of the Safety Proposition being true.

⁸⁷ *Id.* at 9 n.5.

⁸⁸ *Id.* at 10.

⁸⁹ UNDERWRITERS LAB’S. ANSI/UL 4600 STANDARD FOR SAFETY FOR THE EVALUATION OF AUTONOMOUS PRODUCTS (3rd ed. 2023). Available at: <https://www.shopulstandards.com/ProductDetail.aspx?productid=UL4600>.

⁹⁰ Reinvent S-4, *supra* note 9, at 252 (emphasis supplied).

Moving “toward commercialization more safely and quickly” refers to a testing protocol in development by Aurora which allows virtual testing of AV technology using particularly efficient simulations which reduces the need for highway testing. Highway testing places drivers, pedestrians, cyclists, and others at risk of death and injury, making a reduction in highway testing preferable. That is what Aurora means by “more safely and quickly.” That this testing method is superior for safety should suggest just how important gauging the truth of the Safety Proposition is for the protection of the public prior to an actual deployment.⁹¹

If isolated testing of an AV on public highways is dangerous, how much more so is deploying AVs at scale on public highways when the Safety Proposition is false? “Testing” safety may not compare to “deployment” safety because safe testing deployment with a human driver back-up may not be safe enough to deploy without the backup. Testing safety assumes the human back-up will save the day in case of computer driver error. This is one reason driverless “testing” is problematic.

Aurora has the laudable goal to develop AV technology that “can safely and reliably navigate its environment.”⁹² But this is not necessarily the same goal as developing an AV technology where computer drivers are superior to human drivers because Aurora does not indicate how “safely and reliably navigate its environment” ought to be measured. Would a standard of an 18-year-old average male high school student with a new driver’s license suffice?⁹³ If we allow these young drivers on the road when we know this group is riskier than the average driver,⁹⁴ then perhaps an AV with a similar safety profile ought to be permitted.

⁹¹ The industry has failed to follow its own consensus standard which recommends use of back-up human drivers. SAE INT’L. (2020). SAFETY-RELEVANT GUIDANCE FOR ON-ROAD TESTING OF PROTOTYPE AUTOMATED DRIVING SYSTEM (ADS)-OPERATED VEHICLES J3018_202012 (2020), https://www.sae.org/standards/content/j3018_202012/.

⁹² Reinvent S-4 at 252.

⁹³ See, e.g., Noah J. Goodall, *Potential Crash Rate Benchmarks for Automated Vehicles*, 2675 J. TRANS. RESEARCH BD. 31, 38 (July 28, 2021), <https://journals.sagepub.com/doi/10.1177/03611981211009878> (noting that there is no consensus as to how safe an AV must be before deployment).

⁹⁴ DAVID S. LOUGHRAN, SETH A. SEABURY & LAURA ZAKARAS, *WHAT RISKS DO OLDER DRIVERS POSE TO TRAFFIC SAFETY?* (RAND Corp. 2007) (noting that drivers 65 and older are 16 percent likelier than adult drivers (those 25–64 years old) to cause an accident, and they pose much less risk to the public than do drivers under 25, who are 188 percent likelier than adult drivers to cause an accident).

V. THE PIVOT TO “EDUCATION”

The Registration Statements hint at a strategy to manage public expectations about safety⁹⁵ by confirming the goal of making roads safer but also indicating that Aurora will educate the public by teaching the “benefits” of AV technology.

Achieving our mission—*delivering the benefits of self-driving technology safely, quickly, and broadly*—is how we aim to make a positive impact in communities. *We strive to revolutionize transportation by making roads safer*, providing better services for people who currently have difficulty accessing transportation, freeing up time during commutes, and helping goods more efficiently reach those who need them. . . .

As part of our commitment to work with communities where Aurora has a presence and *educate them on the benefits of self-driving technology*, we partner with local organizations that help bridge the digital divide and promote STEM education.⁹⁶

The passage assumes AV technology will have benefits, though the reference to “strive” makes clear that any “benefits” are merely potential. Yet the passage mentioning “benefits” reads as if there is no doubt that AV technology already has demonstrated these benefits. The generic idea is that eliminating a variety of human driver errors that cause traffic accidents makes the computer driver safer. This hypothesis might prove true, but one could only reach that conclusion after understanding the separate risks presented by the computer driver. Currently we do not know whether a minimally safe AV is possible with current and near-term feasible technology.

The Registration Statements do not disclose the true extent of this education effort. In fact, there is a 501(c)(3) called “Partners for Automated Vehicle Education” (sometimes referred to as “PAVE”) which is dedicated to this education project. Aurora is one of 63 members of PAVE. This tax-exempt organization has a single stated goal:

⁹⁵ See Reinvent S-4, *supra* note 9, at 249 (“Corporate Social Responsibilities and Sustainability”).

⁹⁶ *Id.* at 249 (emphasis supplied).

To bring the conversation about automated vehicles (AVs) to the public so everyone can play a role in shaping our future. PAVE's goal is purely educational—we don't advocate for a particular technology or specific public policies. Our members believe that we can only achieve the potential benefits of driverless technology if the public and policymakers *know the honest facts*. PAVE wants to raise public awareness of both what is on the roads today and what is possible for the future.⁹⁷

The AV industry faces a skeptical public. PAVE has done polling which indicates the public has significant doubts about the safety of AV technology.⁹⁸ PAVE's goal appears to be to “sell” the public on the idea that the benefits of AV technology outweigh the risks. Benefits exist if it could be shown to a reasonable certainty that computer drivers were in fact safer than human drivers in the full range of highway conditions. The AV industry would not need a separate education effort if the current science supported this conclusion—but it does not yet do so.

The education campaign may, however, increase shareholder value in AV companies. And a public “educated” through advocacy might more readily accept a short-term reduction in safety to reap potentially substantial longer-term benefits. This possibility presents a public policy debate worth having. But engaging in education efforts about potential benefits without disclosing the deployment standard seems to break trust, because a commitment to deploy when the AV technology is “sufficiently safe” has no content without an articulated standard.

The truth of the Safety Proposition is unknown and will remain so at least until extensive further testing. Extolling the benefits of AV technology as “fact” is premature and dissembling.

VI. CHANCE SET-UPS

To understand the complexity of the AV Problem, consider some toy models⁹⁹ of highway safety problems. Following Ian Hacking, I call

⁹⁷ See PAVE, *About*, <https://pavecampaign.org/about/> (last visited Aug. 26, 2023) (emphasis supplied).

⁹⁸ PAVE Poll: Americans Wary of AVs But Say Education and Experience with Technology Can Build Trust, PAVE (reporting results from a poll taken between Feb. 27 and Mar. 5, 2020), <https://pavecampaign.org/pave-poll-americans-wary-of-avs-but-say-education-and-experience-with-technology-can-build-trust/> (last visited Aug. 26, 2023).

⁹⁹ On toy models, see DANIEL C. DENNETT, *INTUITION PUMPS AND OTHER TOOLS FOR THINKING* (W. W. Norton Company 2014).

these “chance set-ups.” “[T]he frequency in the long run of accidents on a stretch of highway seems to be a property of, in part, the road and those who drive upon it. We have no general name in English for this sort of thing. I shall use ‘chance set-up.’”¹⁰⁰

Consider two chance set-ups. Urn number one (U1) contains 5 black balls and 95 white balls. Urn number two (U2) contains 1 black ball and 99 white balls. The black balls indicate an accident. U1 represents the chance of an accident on a stretch of road with human drivers. U2 is a possible world with computer drivers on the same stretch of road¹⁰¹ with a lower chance of accident. The urn scenarios transparently reveal the relative accident frequencies by revealing the relative mix of black and white balls. In the real world we cannot look inside the urns to determine the relative frequency of accidents on a stretch of highway because the chance set-up of a real stretch of highway does not wear relative frequencies on it sleeve like a coin in a coin flip or a die in a roll of dice.¹⁰²

In the real world we infer the relative frequency of accidents in the U1 scenario by examining actual “trials” based on data such as that collected by the Bureau of Transportation Statistics from accidents which we might use to form a reasonable belief about relative accident frequencies in different environments with human drivers.¹⁰³ The AV companies anticipate that computer drivers will have a lower relative frequency of accidents, like that reflected in U2, but as of now they have nothing to prove such a fact. This explains the pivot to an “education” campaign.

Eliminating human errors (such as texting, drunk driving and falling asleep) might produce this U2 world, but we cannot infer that outcome by simply replacing human drivers with computer drivers. We *do know* that the computer driver will not make many of the same errors as the human driver because the computer driver does not drink and drive, text or fall asleep. But we *do not know* the full range of types and

¹⁰⁰ IAN HACKING & JAN-WILLEM ROMELIJN, *LOGIC OF STATISTICAL INFERENCE* 12-24 (Cambridge Phil. Classics 2016).

¹⁰¹ The stretch of road needs to be the same for each urn so that the comparison of accident experience is apples to apples by using a single operational design domain for both the machine driver and the human drivers.

¹⁰² Using a single stretch of road is an idealized scenario that is unlikely to occur in real world comparisons. However, real world comparisons ought to strive for comparability—such as comparing urban environments to urban environments, and rural environment to rural environments, divided highway experiences to divided highway experiences, etc.

¹⁰³ U.S. DEPT. OF TRANS., *Bureau of Transportation Statistics, Transportation Accidents by Mode*, <https://www.bts.gov/content/transportation-accidents-mode> (last visited Aug. 26, 2023).

frequency of errors that a computer driver will make. All we know for sure is that most computer driver mistakes will differ in kind from human driver mistakes, such as mistaking a Buddhist temple for an ostrich or failing to treat workers wearing yellow safety vests as persons (as discussed in the next following section).

In actual practice, AVs will perform in scenarios with ever increasing AV density as the number of AVs deployed on our highways increases: 10% AVs, 50% AVs, 90% AVs. As dynamic systems, accident frequencies may vary based on different densities. It is not an analytic or *a priori* truth that a higher density of AVs always produces a lower accident rate. Thus, inferring anything about the composition of white and black balls in U2, particularly as the mix of vehicles changes, is challenging. In contrast, there is real world data collected by the Bureau of Transportation Statistics from which to infer the composition of U1 to a reasonable approximation.¹⁰⁴

VII. NEURAL NETWORKS, BUDDHIST TEMPLES AND OSTRICHES

A particular concern with AV technology is the *perception function* which evaluates the environment in which an AV is located. Image analysis which allows for object and event detection and response (OEDR) systems to perform the perception function must use neural networks. The Reinvent S-4 refers to “machine learning” to describe aspects of its system relying on neural networks. Examples of neural network failures in the perception function include identification of a Buddhist temple as an ostrich;¹⁰⁵ and, failure to identify persons wearing yellow or lime green safety vests as persons.¹⁰⁶ No one anticipated these problems. Tesla test footage reveals a failure to identify a massive concrete column supporting a monorail system which almost resulted in a crash.¹⁰⁷ And Tesla’s system has mistaken the Moon for a yellow

¹⁰⁴ U.S. Dept. of Trans., NHTSA, *Estimates for 2022 Show Roadway Fatalities Remain Flat After Two Years of Dramatic Increases* (Apr. 20, 2023), <https://www.nhtsa.gov/press-releases/traffic-crash-death-estimates-2022>.

¹⁰⁵ Aakarsh Yelisetty, *How to Fool a Neural Network?*, TOWARDS DATA SCI. (Aug. 8, 2020), <https://towardsdatascience.com/how-to-fool-a-neural-network-958ba5d82d8a>.

¹⁰⁶ *Id.* (This example was given to the author while receiving technology demonstrations at the Georgia Institute of Technology computer science laboratory.).

¹⁰⁷ Tim Levin, *From Swerving into a Median to Narrowly Missing Poles, Videos of Tesla's Latest Full Self-Driving Update Don't Inspire Much Confidence*, BUS. INSIDER (July 14, 2021, 4:30 PM), <https://www.businessinsider.com/tesla-fsd-full-self-driving-videos-flaws-glitches-2021-7>.

caution light¹⁰⁸ and a Burger King advertisement as a stop sign.¹⁰⁹ A Cruise robo-taxi failed to identify wet cement.¹¹⁰ Consider further challenges for object and event detection and response in environments with children in Halloween costumes or persons wearing scarves and other face coverings.

Part of safety consists not only in making proper object identification, but also in predicting future movements. Consider training needed to predict the pedestrian behavior of drunken sports fans leaving a stadium after an event. Human drivers recognize the risks posed by various situations—such as “these pedestrians are inebriated” or “when a ball crosses a street, children may follow.” As explained to the author during a visit to the computer science lab at Georgia Tech, a typical unimpaired human driver also quickly recognizes mistakes and promptly corrects for them.

Computer drivers do not yet appear particularly good at self-correction. One reason for this is that a human driver understands clues from life experiences—such as a pedestrian on the left is more likely to run across a street to catch a bus on the right than a pedestrian on the left when no bus is present. A human driver might slow down as a prudential matter in a 25-mph zone to 10 miles per hour if on a street with many pedestrians at certain times of the day, or due to a special event such as a wedding reception outside a church. Absent specific training for special situations, the AV might continue at the posted speed limit because it lacks these life experiences.

Images and environments like these are particularly concerning because it is impossible to anticipate and train for all of them. The problem is compounded because neural networks simply exhibit unpredictable behavior.¹¹¹ In the Buddhist temple/ostrich case, the failed identification resulted from adding a small bit of noise to the image of the temple.¹¹² A human eye would not have recognized a difference in the image.

¹⁰⁸ See Jay Ramey, *Tesla’s Full Self Driving System Mistakes the Moon(!) for Yellow Traffic Light*, MSN (July 23, 2021), <https://www.msn.com/en-us/autos/news/tesla-s-full-self-driving-system-mistakes-the-moon-for-yellow-traffic-light/ar-AAMuszU>.

¹⁰⁹ See Levin, *supra* note 107.

¹¹⁰ Joshua Bote, *Cruise Vehicle Gets Stuck in Wet Concrete While Driving in San Francisco*, SFGATE.COM (Aug. 15, 2023), <https://www.sfgate.com/tech/article/cruise-stuck-wet-concrete-sf-18297946.php>.

¹¹¹ Heather M. Ruff, *The Folly of Trolleys: Ethical Challenges and Autonomous Vehicles*, BROOKINGS (Dec. 17, 2018), <https://www.brookings.edu/research/the-folly-of-trolleys-ethical-challenges-and-autonomous-vehicles/>.

¹¹² Yelisetty, *supra* note 105.

Traditional engineering safety analysis assumes that small visual differences will not adversely affect the performance of a product because a typical physical product has a range of acceptable tolerances. Neural networks differ, and this difference changes the safety approach engineers must use. Even though a neural network is a deterministic system, it is hard for humans to analyze its function using existing technologies; its operation is not transparent like the operation of an algorithm. The change of a single pixel in an image can mean the difference between a proper identification and a failure.

Mistakes in a neural network are corrected after the fact by additional training using exposure to large data banks of images and dynamic scenes. Retraining might improve a neural network on one dimension while making it worse on another (as explained to the author at Georgia Tech).¹¹³ One cannot perform “surgery” on a neural network. It is difficult and expensive to create these data banks and there are only a few of them¹¹⁴ used to create the core functionality of a perception function for autonomous driving. Developers supplement data banks containing actual scenes with computer simulations. Recent research focuses on methods to test neural networks for accuracy to reduce mistakes like occurred in the Buddhist temple/ostrich mix up.¹¹⁵

The Reinvent S-4 discloses that Aurora uses non-machine learning based algorithms to perform most of the work in the motion controller for its system (only relying on machine learning systems for certain small adjustments). The Reinvent S-4 refers to an algorithm as an “engineered solution.” Algorithms exhibit predictable behavior and can be debugged. Risk control for a neural network used to supplement an algorithm in a motion controller might engage the neural network for motion control functions at lower speeds or only within a narrow window

¹¹³ See generally Porter, *supra* note 85.

¹¹⁴ See, e.g., KITTI-360, *Kitti Vision Benchmark Suite from Germany*, <http://www.cvlibs.net/datasets/kitti/> (last visited Aug. 25, 2023). Tesla collects information from its customers’ driving experiences and other developers use information from pilot projects to supplement these databases. See Brad Templeton, *Cruise Can Test Rider-Only Robotaxi Rides in California for No Fee*, FORBES (June 6, 2021), <https://www.forbes.com/sites/bradtempleton/2021/06/06/cruise-can-test-rider-only-robotaxi-rides-in-california-for-no-fee>. See also Ryan Duffy, *We Spent Two Days Riding Waymo’s Driverless Taxis and It Felt Surprisingly Normal*, EMERGING TECH BREW (July 16, 2021), <https://www.morningbrew.com/emerging-tech/stories/2021/07/16/spent-two-days-riding-waymos-driverless-taxis-felt-surprisingly-normal>.

¹¹⁵ Chuchu Fan, *Formal Methods for Safe Autonomy: Data-Driven Verification, Synthesis, and Applications*, (2019) (Ph.D. dissertation, Univ. of Illinois, Urbana-Champaign), <https://www.ideals.illinois.edu/bitstream/handle/2142/106202/FAN-DISSERTATION-2019.pdf>.

set by the algorithm, relegating the neural network to fine tuning and ride smoothing, without the ability to direct movement to a zone outside the safety envelope around the AV.

Despite Aurora’s apparently prudent steps for its motion controller, a neural network necessarily performs the critical perception function on which the system makes motion control decisions. In addition to problems with neural networks, many challenges remain to get the algorithmic part of the AV system to work well. A vast amount of uncertainty remains.

VIII. IMPORTANCE OF ALTERNATE TESTING METHODS

Alternate testing and validation techniques to assess the efficacy of training methods to correct perception mistakes are critical to getting AV technology ready for deployment.¹¹⁶ The Reinvent S-4 clearly indicates that Aurora understands this fact because simulation training forms a key part of its business strategy.¹¹⁷ A study by the Rand Corporation stated:

Given that current traffic fatalities and injuries are rare events compared with vehicle miles traveled . . . fully autonomous vehicles would have to be driven hundreds of millions of miles and sometimes hundreds of billions of miles to demonstrate their safety in terms of fatalities and injuries. Under even aggressive testing assumptions, existing fleets would take tens and sometimes hundreds of years to drive these miles—an impossible proposition if the aim is to demonstrate performance prior to releasing them for consumer use.¹¹⁸

¹¹⁶ Some wrongly think that this question of safety has been decided. See McGee, *supra* note 1 (noting “[t]hat dispute is over. Google’s project, now branded Waymo, has experienced only minor incidents.”). The AV industry cultivates this misperception that the safety improvements promised for AVs are a fact, not a projection.

¹¹⁷ Reinvent S-4 at 242 (noting that “Aurora has invested significantly in virtual testing at a time when much of the self-driving industry was focused on real world mileage accumulation”). The efficacy of virtual testing also needs validation.

¹¹⁸ NIDHI KALRA & SUSAN M. PADDOCK, RAND CORP., DRIVING TO SAFETY: HOW MANY MILES OF DRIVING WOULD IT TAKE TO DEMONSTRATE AUTONOMOUS VEHICLE RELIABILITY?, (2016), https://www.rand.org/content/dam/rand/pubs/research_reports/RR1400/RR1478/RAND_RR1478.pdf.

The Rand report shows that alternative methods must supplement road testing to demonstrate safety.¹¹⁹ AV companies must meet this technical challenge of alternative testing before an AV company might make a responsible decision to deploy an AV system. Actual highway testing must follow to validate the efficacy of the simulations. Aurora notes: “The successful development of our self-driving systems and related technology involves many challenges and uncertainties, including: . . . *successfully completing system testing, validation, and safety approvals.*”¹²⁰

To demonstrate safety of an AV system, AV companies supplement simulations with metrics from actual road testing, such as miles driven without an incident or miles driven without human intervention.¹²¹ Given existing technology, such data does not conclusively show that a computer driver is safer than a human driver from an accident severity perspective because, among other things, we do not see what would have happened in the absence of human intervention.¹²² The concern remains, however, that aggressive financial timetables set by investors may not allow sufficient time for continued testing and analysis to better assess the truth of the Safety Proposition.

IX. STRUCTURES TO PROMOTE RESPONSIBLE DECISION MAKING

The Registration Statements’ disclosure creates doubt about whether Aurora has the internal procedures in place to protect against the risk of a bad decision against the public interest. A business model that “set[s] outrageous goals” might create a greater than normal temptation when confronting an ethical dilemma. The Registration Statements describe Aurora’s generic commitment to safety but does not

¹¹⁹ Fan, *supra* note 115.

¹²⁰ Reinvent S-4, *supra* note 9, at 29 (emphasis supplied).

¹²¹ See, e.g., Matthew Schwall, et. al, *Waymo Public Road Safety Performance Data*, WAYMO LLC (2020), <https://storage.googleapis.com/sdc-prod/v1/safety-report/Waymo-Public-Road-Safety-Performance-Data.pdf>. The California Public Utilities Commission resolutions of August 10, 2023, expanding the scope of robotaxi operations in that state identified only 790,000 vehicle miles traveled (VMT) by Cruise (Resolution TL-19145) and 1.7 million VMT by Waymo (Resolution TL-19144).

¹²² Waymo claims to learn what would have happened if a back-up safety driver had not intervened through its computer simulations. The small incremental bit of knowledge gained by observing actual crash results in the absence of an intervention is the only argument for conducting unmanned testing without a safety driver. Until the point of intervention, the information learned from a test with a backup safety driver is identical to the information learned from an unmanned test. The “look Mom, no hands” approach of conducting tests without human safety drivers is done primarily to impress investors not to gather useful scientific information. See Widen, *supra* note 4.

situate this commitment in any ethical framework other than to disclose that Aurora will “do the right thing.” Aurora and other AV companies might adopt IEEE 7000 as a framework specified by engineers for engineers. Aurora might use other corporate governance structures as well.

On page 269 of the Reinvent S-4 one finds the first use of the word “ethics.” It states that Aurora will have a code of ethics which will be available on its website.¹²³ Aurora’s posted corporate codes relate to conflicts of interest, condemnation of discrimination, creating a safe and respectful working environment and similar matters. They do not specifically address an approach to safety. Aurora has formed a Safety Advisory Board.¹²⁴ However, there has been no public commitment to grant that board veto power on a deployment decision the safety board considers insufficiently safe.¹²⁵

To foster trust, Aurora might craft a different sort of ethics code which included more details, specifically addressed to safety issues and the decision process for deployment of AV technology. One common technique used to address concerns over potentially tainted decision making is to form a special committee of the board of directors. Such a committee might consist of three independent directors with technical backgrounds and no financial interest in Aurora, with power and funding to engage outside safety experts, much as a special committee in a going private transaction consults valuation experts to provide a fairness opinion for a sale of the company when management insiders have a conflict of interest.¹²⁶

¹²³ Aurora has posted these materials. AURORA’S CODE OF CONDUCT AND ETHICS and AURORA’S CORPORATE GOVERNANCE GUIDELINES available at *Investor Relations, Corporate Governance Documents*, AURORA, <https://ir.aurora.tech/corporate-governance/governance-documents> (last visited Aug. 25, 2023) (follow “Code of Conduct and Ethics” hyperlink and “Corporate Governance Guidelines” hyperlink).

¹²⁴ See Nat Beuse, VP of Safety, *Our Updated Safety Report and First-ever Safety Advisory Board*, AURORA (June 2, 2021), <https://aurora.tech/blog/aurora-shares-safety-report> (last visited Aug. 25, 2023),

¹²⁵ *Id.* (failing to describe the powers of the Safety Advisory Board). The safety advisory board is not referenced in any of: the Certificate of Incorporation of Aurora Innovation, Inc. (effective Nov. 3, 2021), filing on Form 8-K (Exhibit 3.1) (Nov. 4, 2020), <https://www.sec.gov/Archives/edgar/data/0001828108/000119312521319524/d249044dex31.htm>, or the Bylaws of Aurora Innovation, Inc. (adopted on Nov. 3, 2021), filing on Form 8-K (Exhibit 3.2) (Nov. 4, 2020), <https://www.sec.gov/Archives/edgar/data/0001828108/000119312521319524/d249044dex32.htm>.

¹²⁶ Another approach would be to expressly follow the development procedures specified in IEEE 7000.

The code of ethics might detail procedures for a special “safety” committee, requiring management to present a safety case for deployment, including detailed specifications of the appropriate metrics and the data which justifying deployment as safe to a reasonable certainty. The safety case should disclose any counterarguments against deployment and how those are addressed. Importantly, the safety case should expressly identify the level of safety deemed sufficient and the moral grounds for deployment (e.g., whether a utilitarian case is justified by losses now, with gains to follow), and how that justification complies with Aurora’s prior statements about deployment standards (ideally the truth of the Safety Proposition to a reasonable certainty). The public needs to know whether Aurora is either applying a “patch your way to perfection” standard or is deploying with no real idea about the truth of the Safety Proposition. This type of special safety committee exceeds the protections provided by Aurora’s current safety advisory board (which is not emphasized in the Reinvent S-4) and appears to have no real power.¹²⁷

We know corporate decision makers often make serious mistakes. Consider Ford’s disastrous design decision for the Pinto where corporate managers made an incorrect cost benefit analysis, the Challenger space shuttle disaster where persons lower in a decision tree with better technical knowledge of risks did not stop a decision to launch, and the recent shameful behavior of Volkswagen in concealing the true emissions performance of its diesel engine powered vehicles.

Even if a corporate ethics code touched on matters such as this, the above measures offer no panacea if Aurora faces an existential crisis because moments of crisis provide opportunities for creative rationalizations. Nevertheless, adopting a decision framework in advance of an exigency, and then sticking to it, may help AV companies like Aurora “do the right thing.”

A strategy to promote moral decisions might further include consideration of the famous “Trolley Problem” from ethics classes. Considering a problem like this has value for educating management and

¹²⁷ Kristen Korosec, *Aurora Brings in Outsiders to Boost Safety Efforts, Public Trust of Driverless Vehicles*, TECHCRUNCH (June 3, 2021), <https://techcrunch.com/2021/06/03/aurora-brings-in-outsiders-to-boost-safety-efforts-public-trust-of-driverless-vehicles/>. It is difficult to see how a committee with no real power can be more than a marketing effort, even if some members “live and breathe tech.”

directors about how to responsibly confront and solve an ethical dilemma.¹²⁸

X. COMPARING THE TROLLEY PROBLEM TO THE AV PROBLEM

A. *The Difference Between a Trolley Case and a Trolley Problem*

The Trolley Problem helps one understand the issues at stake in the AV Problem.

The classic trolley problem goes like this: You see a runaway trolley speeding down the tracks, about to hit and kill five people. You have access to a lever that could switch the trolley to a different track, where a different person would meet an untimely demise. Should you pull the lever and end one life to spare five?¹²⁹

The literature on the ethics of AV design substitutes a machine actor for a human actor, and an AV for the trolley. One scenario asks whether computer software in an AV should function to swerve the AV to hit one pedestrian rather than five. A more troubling variant asks whether the AV ought to hit five homeless persons rather than one investment banker.

These ethical dilemmas each present a trolley *case*, and not a trolley *problem*.¹³⁰ The original trolley *problem*, as framed by Judith Jarvis Thomson,¹³¹ compared the person at the switch with a doctor deciding whether to harvest the organs of one person to save five. The “problem” was to explain why there is universal condemnation of a decision by a doctor to harvest organs from one to save five, yet everyone Thomson polled either thought pulling the switch to sacrifice one to save

¹²⁸ IEEE 7000 suggests that a company developing technology ought to consider substantive ethical theories such as utilitarianism as part of a responsible design process and creates a framework within a technology company to focus on such considerations.

¹²⁹ Karen Hao, *Should a Self-driving Car Kill the Baby or the Grandma? Depends on Where You're From*, MIT TECH. REV. (Oct. 24, 2018), <https://www.technologyreview.com/2018/10/24/139313/a-global-ethics-study-aims-to-help-ai-solve-the-self-driving-trolley-problem/>.

¹³⁰ Rob Lawlor, *The Ethics of Automated Vehicles: Why Self-driving Cars Should not Swerve in Dilemma Cases*, 28 RES PUBLICA 193 (2021), <https://link-springer-com.proxy.library.nd.edu/article/10.1007/s11158-021-09519-y>.

¹³¹ See Thomson, *supra* note 2.

five in a trolley case was permitted or, stronger yet, mandatory (except possibly for John Taurek).¹³²

The original trolley problem asks: Why is it permitted to cause the loss of one life to save five in the trolley case, yet it is not permitted to cause the loss of one life to save five in the transplant case? Several subsidiary questions surface, such as whether there is a moral difference between action and inaction, between doing and allowing, and whether a personal right not to be harmed might take precedence over a general rule to maximize happiness and minimize pain for the greatest number of persons. Philippa Foot, who originally posed the scenarios which Thomson labeled as the “Trolley Problem,” suggested that one might morally decide to pull the lever because it was not done for the purpose of killing the one, but rather was done to save the five. Under the doctrine she called “double effect” it was morally acceptable to move the lever even though harm was foreseen because the harm was not the intended result. This “double effect” intuition conflicts with recently articulated suggestions of principles to guide the training of machine learning components of automated vehicles.¹³³

Philosophers pose trolley problems to compare our intuitions about two different moral dilemmas to investigate why moral intuitions differ in cases which have a surface similarity. It does not matter that a trolley problem poses an artificial question because the problem is a thought experiment.

Those who worry that the ethical dilemma in a trolley case is a live problem that needs solving for AVs fail to understand the purpose of the problem. Those who wrongly perceive a problem that needs solving today, go out and try to solve it. That explains much of the ink spilled over AVs and trolley problems,¹³⁴ and why AV executives (like Aurora’s

¹³² See generally John Taurek, *Should the Numbers Count?*, 6 PHIL. & PUB. AFFAIRS, 293 (1977).

¹³³ A. D’Amato, et al., *Exceptional Driving Principles for Autonomous Vehicles*, 2022 J. L. & MOBILITY 2, 15 (2022) (suggesting that an ADS be programmed to maintain a strict duty of care to each road user). A strict duty of care embodied in traffic codes suggests that one may not violate a duty of care owed to one road user to resolve a conflict with another. *Id.* BMVI (2017) rejects the notion that a calculus of any sort should be used to offset one life against another in an unavoidable crash, especially in its rule number 9. BMVI. (2017). *Ethics Commission: Automated and Connected Driving Report*, FEDERAL MINISTRY OF TRANSPORT AND DIGITAL INFRASTRUCTURE, GERMANY, JUNE 2017. Available at: <https://perma.cc/6UBX-KH5G>.

¹³⁴ See Abby Everett Jacques, *Why the Moral Machine is a Monster, We Robot Conference*. University of Miami School of Law (2019), <https://robots.law.miami.edu/2019/wp-content/uploads/2019/03/MoralMachineMonster.pdf>. Jacques observes “The right question isn’t what would I do if I were forced to choose between swerving and going straight. The right question is what kind of world will I be creating if this is the *rule*.”

CEO, Chris Urmson) have appropriately stated that they do not consider trolley problems.¹³⁵

The real question with AV technology is not how to program an AV to perform in an individual case of constrained choice, nor even to consider the impact if a rule for an individual case were to be generalized. The real question is identification of the moral justification for the decision to deploy AV technology in the first place at the time of deployment at scale.¹³⁶ Rather than worry about the structure of the world under different rules, the first step considers the structure of the world immediately after the deployment decision.

AV companies ignore these sorts of philosophical puzzles because current software does not address ethical dilemmas of this sort nor is software likely to do so for the foreseeable future because of the computing power needed on board an AV to make the necessary calculations (as explained to me at Georgia Tech). Philippa Foot recognized back in 1967 that any real-world decision to “pull the lever” is based on probabilities of collision and harm, and not certainty. Performing probabilistic calculations such as these is immensely complex and might exceed the computing power available on board a vehicle. Additionally, the speed of computation might be a problem due to latency.¹³⁷ The practical concerns of safety and accident prevention predominate AV industry thinking because the primary current industry goal is simply getting vehicles from point A to point B without a collision.¹³⁸

B. Application of a Trolley Case to the AV Problem

The trolley case presents a constrained binary choice which determines how many lives will be lost depending on the choice that is made. That is exactly the choice that AV companies face when they decide whether they ought to deploy AV technology at scale. If the Safety

¹³⁵ Hirsch, *supra* note 19 (reporting a comment of Aurora’s CEO).

¹³⁶ Jacques is correct that the issue is not the outcome of a particular accident case, but she does not focus on the morality of the deployment decision itself, focusing instead on the effect of generalizing the result of individual accident cases.

¹³⁷ Daniel Schwartzberg, *In-Vehicle Data Latency: Fast or Furious*, ELEC. DESIGN (Oct. 30, 2019), <https://www.electronicdesign.com/markets/automotive/article/21808774/invehicle-data-latency-fast-or-furious>.

¹³⁸ AV companies also reject the notion that problems raised by the Moral Machine experiment need a solution. They see little risk that society will make choices based on personal characteristics of potential accident victims, believing that legislation will prevent the parade of horrible choices even if technology catches up to make those choices possible.

Proposition is true, then the decision to deploy is justified in the same manner as one would justify a decision to pull the lever in a trolley case. If the Safety Proposition is not true, then the deployment justification evaporates. Moral certainty that the Safety Proposition is true sustains the justification.

An AV Problem restricted to a choice between urns U₁ and U₂ with certainty about the relative frequency of losses on the stretch of highway resembles the trolley case in many respects. Just as Thomson found that everyone she consulted would at least permit one to pull the lever, almost everyone would favor deploying the AV system if the relative frequencies of loss are known to a certainty.

The difference between the two cases is that the decision in the trolley case involves a decision to cause the death of an identifiable person, whereas the decision in the AV Problem does not. Studies have shown that people tend to value identified lives more than statistical lives.¹³⁹ Moreover, people have a greater psychological aversion to direct physical action which causes harm than to causing harm from a distance.¹⁴⁰

Some have suggested, however, that the engineers who create software to “hit one, not five” in a trolley case are more to blame than a human driver reacting in the pressure of the moment because the software designers act with clear deliberation.¹⁴¹ If this is right, then management making the deployment decision, as well as the software engineers, ought to be subject to the same degree of blame. The software designer in the AV case does not target any specific person, and neither does the AV company’s management when it decides to deploy AVs. Yet in both cases the technology arguably results in “targeting” a different group for harm. The five persons killed in the U₁ scenario almost certainly do not include the one person killed in the U₂ scenario because the circumstances of the accident in a U₁ case differ from those in a U₂ case. This resembles the trolley case where different workers appear on different tracks.¹⁴²

¹³⁹ See, e.g., James K. Hammit & Nicolas Treich, *Statistical vs. Identified Lives in Benefit-Cost Analysis*, 35 J. RISK & UNCERTAINTY 45, 46 (2007).

¹⁴⁰ Lauren Cassani Davis, *Would You Pull the Trolley Switch? Does it Matter? The Lifespan of a Thought Experiment*, THE ATLANTIC (Oct. 9, 2015), <https://www.theatlantic.com/technology/archive/2015/10/trolley-problem-history-psychology-morality-driverless-cars/409732/>.

¹⁴¹ Patrick Lin, *Here's a Terrible Idea: Robot Cars with Adjustable Ethics Settings*, WIRED (Aug. 18, 2014, 6:38 AM), <https://www.wired.com/2014/08/heres-a-terrible-idea-robot-cars-with-adjustable-ethics-settings/>.

¹⁴² AV technology might involve risk shifting away from wealthy individuals with safe vehicles onto more vulnerable road users even if statistics showed a nominal increase

The choice in the AV Problem will have a significant societal impact whereas the answer supplied in a trolley case has low impact because it presents an extremely unlikely individual case. In the AV Problem, the binary choice to deploy or not is the essence of the problem. The problem is how to make a responsible judgment about the frequency in a U2 case where we cannot look inside the urn. The uncertainty in the AV Problem and the nature of action at a distance highlight the moral importance of an unbiased assessment of supporting evidence.

C. Bystanders Versus AV Companies

AV companies are not bystanders in the AV Problem because they create the binary choice. Consider whether the following discussion changes the analysis:

The original trolley problem included several variations. Case 1 had the conductor of the trolley make the decision. Perhaps we would view a trolley conductor as “killing” the five workers if she did not change tracks because her status as the conductor created a duty. But steering the trolley onto the track with one worker kills that worker regardless of a duty.

Placing a good or service in the chain of commerce may create a duty in an AV company. Under some legal regimes it is a crime to sell an unsafe product¹⁴³ and a sale might create tort liability. Though moral answers may differ from legal answers, the presence of both civil and criminal liability suggest that placing an unsafe product in the market is not a moral option. But AV technology may not be defective in a moral sense even if it is not demonstrably better than a human driver (or even a bit worse), if other utilities are considered.

While introducing a middling or ho-hum AV product into the chain of commerce may be morally permissible, making misstatements, false representations, and omissions to game statistics or hide problems are not. The situation calls for candor about AV products and the basis

in safety. A net decrease in overall safety might increase the safety for powerful decision makers and allow technological progress to occur, putting aside the important question of personal profit for those in power and discrimination against low-income persons. *See* Widen, *supra* note 4.

¹⁴³ James Nickel directed me to the United Kingdom laws on consumer safety and a publication from the Office for Product Safety and Standards which require that a manufacturer or importer of products into the UK must demonstrate that the product complies with relevant safety requirements. Office for Product Safety and Standards, *Border controls for non-food products*, GOV.UK (Jan. 5, 2021), <https://www.gov.uk/guidance/border-controls-for-non-food-products> (last visited Aug. 26, 2023).

for a deployment decision so a proper public debate might ensue. The original trolley case does not present any choice about disclosure.

Case 2 removes the possibility of a duty because Case 2 places a bystander next to the switch who has an opportunity to pull the lever rather than a conductor. Without an affirmative duty to act, conventional norms suggest that it is morally permissible for the bystander to do nothing. The AV companies' position differs because they developed the very product that creates the binary choice.

Case 3 places the bystander in a position to push a fat man onto the tracks which, it is postulated, will stop the train. Many people who would pull the lever in Case 2 (or, at least, allow that one might have moral permission to pull the lever), become nervous about Case 3 because the action taken is direct and personal. The AV engineers and management are not in the position of the bystander who pushes the fat man. Their responsibility is less direct because their position in the chain of causation is removed from the scene of the accident. Their product, however, is at the scene of the accident. The fact that action at a distance might make it psychologically easier to decide in favor of deployment simply suggests that even more care should be taken that any utilitarian justification has a genuine basis for calculation.

XI. THE NEXT OF KIN QUESTION

John Taurek described to me many years ago his technique to think about moral decisions.¹⁴⁴ He placed himself in the position of having to explain to the next of kin the role his decision played in the outcome that led to the death of their loved one. If he felt uncomfortable with his explanation, he worried his choice was morally unsound.

Management of AV companies might use this method to examine any decision to deploy AV technology at scale. This makes it easy to see the importance of the truth of the Safety Proposition. Confidence in the truth of the Safety Proposition at least to a reasonable certainty allows one to face the next of kin. Explaining that a loved one died so that future travelers on the highway might be safer or that, on balance, this was the best result because delivery costs went down for Amazon Prime likely do not suffice. Justifying a deployment decision in the absence of sound data supporting reasons to believe that computer drivers are substantially more safe than human drivers appear similarly weak

¹⁴⁴ I studied ethics with John Taurek during the 1979-1980 academic year when he was a visiting professor in the philosophy department at Stanford University.

because it resembles gambling with lives for monetary gain. Reference to the mere potential for elimination of human driver error is not enough.

An early deployment decision might avoid a certain near term unpleasant outcome of company failure, but time dimensions may distort rational preferences.¹⁴⁵ The difficult conversation with the next of kin may never happen with a successful deployment so, as a matter of psychology, there is a tendency to postpone unpleasantness and hope for the best.¹⁴⁶ The better moral approach is not to view the potential harm from increased accidents as distant and uncertain compared with the certain harm of immediate financial failure. Rather, before a deployment decision made with inconclusive information management should ask, “How do we personally handle the difficult discussion with the next of kin.” AV development team members might consider whether they would volunteer to be a victim as a technique to assure the problem is approached with the requisite level of moral maturity.¹⁴⁷

Considering past famous accident cases may help personalize the decision. The airplane crash that killed Knute Rockne exposed airline safety as deficient. Regulatory and design changes followed, such as detailed inspections and the use of aluminum framing, rather than wood components.

The Takata airbag recall¹⁴⁸ in which nineteen Americans died from a defective product is instructive. After a small number of airbags exploded on deployment, Takata ended up in bankruptcy.¹⁴⁹ No argument that Takata airbags saved more lives than were lost in accidents passed muster. Takata provides a cautionary tale about blind application of utilitarian justifications.

The response to risks, however, ought not to lead one to stick one’s head in the sand, foregoing air travel and airbags. No new endeavor can practically eliminate all risks. The delicate moral balance considers how to take advantage of technology’s benefits while minimizing risk in a morally responsible way to allow a justification when loss occurs. The decision might become clearer (or at least easier to live with) if an AV company agreed to accept full financial responsibility for loss caused by

¹⁴⁵ See, e.g., Ted O’Donoghue & Matthew Rabin, *Doing it Now or Later*, 89 AM. ECON REV. 1, 103 (1999).

¹⁴⁶ The inclination to delay may be present even in the absence of the moral hazard and uncertainty of outcome.

¹⁴⁷ IEEE 7000 provides an explicit framework within which discussions of this sort might occur.

¹⁴⁸ *Takata Recall Spotlight*, NHTSA, <https://www.nhtsa.gov/equipment/takata-recall-spotlight>.

¹⁴⁹ Tom Hals, Judge approves Takata’s U.S. bankruptcy plan, REUTERS (Feb. 17, 2018), <https://www.reuters.com/article/us-takata-bankruptcy-ruling-idUSKCN1G10SW>.

operation of its AVs.¹⁵⁰ Whereas the moral status of a contrary decision allowing financial responsibility for AV accidents to remain vague seems to speak for itself. That is Aurora's burden, and indeed the burden for the entire AV industry.

Because the justification to deploy AV technology ends up looking a lot like a justification for pulling the lever in a trolley case, AV companies ought to study philosophical puzzles to understand moral justifications and to practice making difficult ethical decisions. That is the important bridge between a fringe philosophy puzzle and the most important question facing the AV industry.

CONCLUSION

Aurora and every other AV company face a problem like a trolley case—one with significant near-term public policy implications. The Registration Statements do not disclose the gravity of the decision nor the implications for public welfare. The consequences of the deployment decision dwarf the ethics of AV design in an individual accident case. This explains the importance for management, directors, and engineers to have a working understanding of the logic and language of ethical discourse.

The way forward has numerous facets. The Registration Statements studied in this essay resemble the kind of dissembling disclosure criticized in the context of financial disclosure of derivatives and other financial products.¹⁵¹ Dancing around the applicable principle for the deployment decision, coupled with downplaying the extent and reasons for the public relations campaign, are just other versions of hiding the truth while letting it appear—like Odysseus appearing to the

¹⁵⁰ Accepting financial responsibility is only a partial solution. One does not get a moral license to harm people simply because one can afford to pay for the consequences. But accepting full financial responsibility, coupled with taking all reasonable efforts to avoid loss, may permit proceeding in the face of uncertainty in a morally acceptable way. Those conditions might allow one to face the next of kin. In Vienna, Austria at "The Autonomous" conference on Sept. 14, 2023, I expanded on this idea in a panel discussion on automated vehicle regulation suggesting that the near-term public relations posture should focus on responsibility in confronting uncertainties while working toward a safer future rather than attempting to demonstrate the truth of the safety proposition on thin evidence using lagging measures. See Junko Yoshida, *When an AV Has the Wheel, Who's Driving?*, THE OJO-YOSHIDA REPORT (Sept. 18, 2023) (describing panel discussion), <https://ojoyoshidareport.com/when-an-av-has-the-wheel-whos-driving/>?

¹⁵¹ William H. Widen, *The Arbitrage of Truth: Combating Dissembling Disclosure, Derivatives, and the Ethic of Technical Compliance*, 66 U. MIAMI L. REV. 393 (2012).

suitors while in a disguise.¹⁵² This is a common disclosure technique in SEC filings for which the poster child is Enron’s footnote disclosure of the Raptor transactions.¹⁵³

Disclosure by indirection often suffices to avoid liability under the securities laws. However, the potential for a broad adverse impact on social welfare in the case of a premature AV deployment decision present considerations different in kind from those typical in SEC disclosure. The question is not solely one of whether investors lose money. Uninvolved members of the public are at risk who made no investment decision. And we all lose if a premature deployment turns the public against a potentially very valuable technology with both personal and national security implications.

For this reason, the AV companies should build trust by announcing the details of their standard for deployment, adopt a fulsome ethics code and form a special board-level safety committee to judge the merits of a safety case. These steps should accompany a decision to implement IEEE 7000. Defending a safety case is more than explaining the potential benefits of AV technology and noting that computer drivers do not text, drink or fall asleep. Rather, there must be a demonstration to a reasonable certainty in a safety case justifying the belief that benefits will immediately follow deployment. The alternative is to deploy on a “harm now, benefits later” justification which raises both moral questions and public relations problems. Properly explained, the public might even allow for a harm now, benefits later deployment if accompanied by AV company acceptance of full financial responsibility for accidents. One cannot tell how the public might react to that package while working to hide the ball.

Review of a conference from 2005 in the wake of the Enron scandal¹⁵⁴ illustrates some of the challenges any proposal to use special committees and other procedures will face, of which I mention a few.¹⁵⁵ Speaking of Enron and the role of directors at that conference, Judge Strine noted, “I think what happened was everybody thought they were geniuses and they thought they were on a board of a company that was managed by geniuses.” Outside directors doing a part-time job can do

¹⁵² MARTIN HEIDEGGER, *PARMENIDES* 27-28 (Indiana Univ. Press 1998). Dissembling “lets something appear otherwise than it is ‘in truth,’ dissembling also unveils and hence is a kind of disclosure.” *Id.* at 44.

¹⁵³ William H. Widen, *Enron at the Margin*, 58 *BUS. LAW.* 961 (2003).

¹⁵⁴ Margaret M. Blair & William W. Bratton, eds., *RESTORING TRUST IN AMERICA’S BUSINESS INSTITUTIONS*, CONF. PROC. NOVEMBER 6-7, 2003, SLOAN PROJECT ON BUSINESS INSTITUTIONS AT THE GEORGETOWN UNIVERSITY LAW CENTER 2005.

¹⁵⁵ *Id.* Particularly Panel III, *The Changing Role of Corporate Directors* at 61-100.

little more than monitor compliance. They have nowhere near the information that the CEO and other top management possess. Yet using special committees places a huge bet on independent directors. Of necessity, that committee will need to engage the assistance of outside safety consultants. The problem is finding enough independent directors and outside consultants to service the entire AV industry with the required expertise and no conflicts.

Many argue that better corporate governance results when the decision makers, including directors, have “skin in the game” through stock options and investments in the company.¹⁵⁶ Yet in this case investments enhance the risk of a significant moral hazard. The panel discussion at the conference revealed that board members ought to have a strong, in-depth understanding of the industry. Motivating persons to join such boards without proper financial incentives may be difficult if stock ownership and options are not available.

The typical concern that management will focus on the short term and ignore the long-term health of the company differs for a speculative emerging business—merely “surviving to the long term” is the goal. These considerations suggest that meaningful self-regulation for AV companies may be aspirational rather than practical. This, in turn, suggests a role for meaningful government regulation—certainly more than currently exists. One ought to consider these ethical problems raised by AV technology rather than focusing on outcomes in uncommon individual accident situations resembling a Trolley Case. Deciding whether AV programming ought to save junior by hitting grandma should rank low on our list of near-term priorities.

Given the stakes for public safety, we ought to use the threat of securities law liability as one motivation to elicit disclosure of the standard for deployment of AV technology so that a proper public debate might take place. The industry’s “education” efforts appear motivated by a desire to silence this debate rather than promote it. If the truth of the Safety Proposition is either, in fact or in principle, unknowable, even to a reasonable certainty at the time of deployment, then this debate is an even more important debate to have. We ought not decide an issue of

¹⁵⁶ This is the problem created when ownership is separated from management as occurs in most public companies, as originally discussed in ADOLF BERLE AND GARDINER MEANS, *THE MODERN CORPORATION AND PRIVATE PROPERTY* (1933). Requiring a minimum ownership percentage in the company, or granting options in company stock, is an attempt to align the interests of management with the interests of shareholders to avoid the moral hazard created when ownership is separated from management. See generally George J. Stigler & Claire Friedland, *The Literature of Economics: The Case of Berle and Means*, 26 *J.L. & ECON.* 237 (1983).

this importance for public welfare by delegating its resolution to a small group of persons with a financial interest in the outcome. If the trajectory does not change to include a candid discussion of the most important ethics decision facing the AV industry, that is the road we will find ourselves traveling upon.¹⁵⁷ We should not wait to have this discussion only after we experience an increase in fatalities as AV technology deploys, AV companies fail and lawsuits pile up for securities law disclosure violations.

POSTSCRIPT

Since the original posting of an early version of this essay in 2021, the quality of public discussion about the ethics of automated vehicle design, testing, and deployment has not improved. Indeed, in context, the situation has gotten worse. I have seen this firsthand at public hearings devoted to automated driving technology.

Troubling incidents continue to highlight the immaturity of the automated driving systems in AVs and these incidents multiply in tandem with the expansion of testing and pilot programs using Level 4 vehicles without human supervision. Nowhere is this immaturity more apparent than in San Francisco following California Public Utilities Commission (CPUC) approval for GM’s Cruise and Alphabet’s Waymo to expand pilot robotaxi operations given on August 10, 2023—a move vigorously opposed by the city’s first responders.

Only days after CPUC approval, San Francisco experienced: traffic jams involving up to 10 robotaxis on a single street; a robotaxi collision with a firetruck involving injuries; a robotaxi collision with a boom truck; a robotaxi getting stuck in wet cement; and more. California does not stand alone as witness to public road operation of a technology not yet ready for prime time. In the same month as CPUC approval, an automated public bus in Orlando had an accident on its second day in service.

Plenty of incidents even prior to CPUC approval should have motivated AV companies to engage in self-reflection and change their approach to public relations and the possibility of limited local regulation. Instead, we find full page newspaper ads proclaiming that humans are terrible drivers. The industry’s rhetoric trajectory remains

¹⁵⁷ Indeed, this was the situation in Pennsylvania with SB 965 in which the AV companies engaged in a process Philip Koopman and I called “autonomandering” in which companies co-op industry friendly legislators (generally from rural and suburban areas) to foist lax AV laws and regulations on urban areas. See Widen & Koopman, *supra* note 11.

the unsupportable claim that “AVs are already saving lives.” Any opposition to industry’s legislative and regulatory wishes amounts to killing people in this dishonest narrative.

In an honest and responsible approach, the AV industry might make the following points. 1. We do not yet know whether AV deployments will save lives. We are hopeful that safety improvements will follow deployments. 2. In the meantime, we are proud that during early operation of series production Level 3 vehicles and commercial operation of Level 4 robotaxis, no fatalities have occurred. 3. In the near to intermediate term, we do not expect AVs to be perfect (and maybe never). Incidents of performance failures have occurred and will continue to occur. We promptly analyze these incidents, making changes to our systems as appropriate. 4. We promote the safety of our AVs by following industry consensus safety standards in the design and testing of our ADS. 5. We are confident in the performance of our products and accept that our AVs each owe a duty of care to other road users and bystanders. We accept responsibility for any breach of this duty of care just as a human driver has responsibility for a breach of the duty of care imposed by law on human driver behavior.

Honesty and candor can help develop public trust in AV technology. Maintaining trust is important to gain public acceptance of the technology. Without public acceptance it will be very difficult to realize any benefits which AV deployments might provide. At present, it is unknown how broadly AVs might be deployed and the roles they might perform in the transportation system (in an economically viable way). We would not want to lose the benefits of AVs for long-haul trucking or transportation to airports on established routes, for example, because deployments in overly broad and optimistic operational design domains resulted in serious accidents that turned the public against automated driving technology more generally.