Routledge
Taylor & Francis Group

Jurisprudence
An International Journal of Legal and Political Thought

ISSN: 2040-3313 (Print) 2040-3321 (Online) Journal homepage: www.tandfonline.com/journals/rjpn20

Enter Technico: a technically informed perspective on moral dilemmas in autonomous vehicle policy

Jan Hölzer

To cite this article: Jan Hölzer (22 Dec 2025): Enter Technico: a technically informed perspective on moral dilemmas in autonomous vehicle policy, Jurisprudence, DOI: 10.1080/20403313.2025.2576959

To link to this article: https://doi.org/10.1080/20403313.2025.2576959

Published online: 22 Dec 2025.

JURISPRUDENCE
https://doi.org/10.1080/20403313.2025.2576959

Routledge
Taylor & Francis Group

REVIEW ARTICLE

OPEN ACCESS Check for updates

Enter Technico: a technically informed perspective on moral dilemmas in autonomous vehicle policy

Jan Hölzer

Institute for Ethics in Technology, Hamburg University of Technology, Hamburg, Germany

In their book 'Moral Dilemmas Involving Self-Driving Cars: How to Regulate Them and Why Your Opinion Matters', Norbert Paulo and Lando Kirchmair present us with a hard pill to swallow: 'Like it or not, we the people have to decide who to save and who to kill.'[1] Having a self-driving car make this decision seems frightening. This uneasy feeling is why accidents involving self-driving cars attract a lot of media attention. Ideas on how to handle these situations have been the subject of numerous research papers.[2] The resulting academic discourse drew attention to the ethics in technology and the emerging field of autonomous vehicle ethics.[3]

Among other work on the ethics of autonomous vehicles, Paulo and Kirchmair's book stands out for their exemplary inclusive and interdisciplinary approach. It presents a series of philosophical questions concerning moral dilemmas involving self-driving cars, contrasts them with empirical data on public opinion and makes a compelling case for why such preferences should be taken into account when determining a justified course of action. The authors are convinced that '[m]orality is not to be found in the "ivory tower" of ethical theorizing; nor is it simply to be found in what (the majority of) laypeople think is right'.[4] According to them, it is necessary to find 'sensible ways of accounting for both public morality and traditional moral theory'.[5]

In this article, I do not seek to fundamentally challenge Paulo and Kirchmair's analysis but to build on it. To achieve this goal, I focus on a few of the most significant parts of the book, offering both a critical discussion of their arguments and a perspective grounded in technical considerations. I argue that without integrating this perspective, their

CONTACT Jan Hölzer jan.hoelzer@tuhh.de Institute for Ethics in Technology, Hamburg University of Technology Hamburg, Germany.

[1]N Paulo and L Kirchmair, *Moral Dilemmas Involving Self-Driving Cars: How to Regulate Them and Why Your Opinion Matters* (Routledge 2025) ix.

[2]A literature review conducted in 2023 identified that the first respective publication emerged in 2014. See F Poszler and others, 'Applying Ethical Theories to the Decision-Making of Self-Driving Vehicles: A Systematic Review and Integration of the Literature' (2023) 75 Technology in Society 102350.

[3]The current literature concentrates on crash scenarios in which an autonomous vehicle must decide how to distribute unavoidable harm, often referred to as trolley cases. For a broader overview extending beyond these scenarios, see, for instance, M Maurer and others (eds), *Autonomous Driving: Technical, Legal and Social Aspects* (Springer 2016) and R Jenkins, D Černý and T Hříbek (eds), *Autonomous Vehicle Ethics: The Trolley Problem and Beyond* (Oxford University Press 2022).

[4]N Paulo and L Kirchmair, *Moral Dilemmas Involving Self-Driving Cars: How to Regulate Them and Why Your Opinion Matters* (Routledge 2025) 74.

[5]Ibid 55.

recommendations risk becoming detached from the regulatory and technical realities they aim to address. To this end, the article proceeds as follows. Section 1 introduces the book's approach and highlights the absence of a technical perspective as an important limitation. Section 2 explores the ethical foundations of autonomous vehicle policy, examining the relevance of philosophical thought experiments. Section 3 analyses the authors' use of empirical data, drawing attention to methodological challenges. Section 4 discusses their effort to integrate moral theory, public opinion, and human rights into a method for informing policymakers. Section 5 addresses how Paulo and Kirchmair's policy recommendations translate into regulatory practice, ultimately arguing that braking – rather than evasive action – should be the default in crash scenarios. Throughout these sections, I propose a set of principles aimed at guiding the translation of ethical reasoning into technical and regulatory practice. In conclusion, Section 6 reflects on how this approach can strengthen the authors' otherwise persuasive proposal and ends with a list of the formulated principles.

1. Setting the stage

Paulo and Kirchmair's account is notable not only for the richness of its content, but also for the distinctive and effective manner of its presentation. The authors employ two fictional characters, the professors *Principia* and *Skeptico*, to simulate a dialogue between two perspectives. Professor Principia is a 'clear-headed'[6], philosophically rigorous thinker, grounded in legal reasoning and deontological ethics. She is sceptical of empirical studies on laypeople's moral preferences, as she believes they tend toward a 'large-scale trivialization of the relevant issues'.[7] In contrast, Professor Skeptico is a critical, empirically informed philosopher who challenges both legal orthodoxy and philosophical dogmatism. While he concedes that empirical studies should not inform policy without reflection, he insists that this 'does not mean that the preferences collected in empirical studies are worthless'.[8]

Building on this exchange, Paulo and Kirchmair offer a short list of recommendations for policymakers on the 'most promising solutions for resolving moral dilemmas involving self-driving cars through regulation'.[9] Among these, they argue that regulation must account for technical realities. Or as they say, 'Regulation needs to talk to code'.[10] This means that policy must be written in a way that can be operationalised within the logic of software systems. Unfortunately, the authors only hint at this challenge and do not take the opportunity to offer an in-depth analysis that could have made an even more significant contribution. Without it, their recommendations risk remaining under-specified and difficult to implement.[11] On two occasions, the authors mention three groups of experts: 'engineers, ethicists, and lawyers'.[12] Yet while the book contains

[6]Ibid 22.
[7]Ibid 50.
[8]Ibid 59.
[9]Ibid 81.
[10]Ibid 83.
[11]See F Poszler and others, 'Applying Ethical Theories to the Decision-Making of Self-Driving Vehicles: A Systematic Review and Integration of the Literature' (2023) 75 Technology in Society 102350, 25. The authors found that 'over time and still to this date, publications with no remark on the technical implementation make up the majority of the publications'.
[12]N Paulo and L Kirchmair, *Moral Dilemmas Involving Self-Driving Cars: How to Regulate Them and Why Your Opinion Matters* (Routledge 2025) 50, 74.

a multitude of ethical and legal insights, it lacks the perspective of an engineer. Another character should join the stage: Enter *Technico*.

Technico is a pragmatic engineer who emphasises real-world feasibility in ethical and legal discussions. Unlike Principia and Skeptico, who represent established academic perspectives, Technico introduces a different kind of expertise. He speaks not in terms of moral theory or survey data, but in terms of technical feasibility and system design. His role is to prevent the discussion from drifting too far into abstraction. By asking how proposed principles can be operationalised within actual systems, he ensures that ethical and legal analysis remains responsive to technical realities.

2. Philosophy meets the road

The book begins with a brief history of car accidents. Paulo and Kirchmair use this historical lens to show that accidents are not exceptions but an inherent part of road traffic. Accordingly, harm reduction emerges as a key ethical concern. The question is whether self-driving cars could be an appropriate means to this end. The authors point out that it is 'very difficult to answer'[13], as there is little data from real situations, because the automated driving systems feasible and legally permitted to date are designed for relatively clear and simple circumstances.[14] This lack of data makes it difficult to conduct reliable studies comparing human and automated driving safety. However, Paulo and Kirchmair argue that there is value in supporting and automating at least some of these tasks. They present a study stating that 'driver-related factors (i.e., error, impairment, fatigue, and distraction) [are] present in almost 90% of crashes'.[15] Given this data, it seems sensible to promote automation, provided that the use of self-driving cars does not create new, more serious problems.

This focus on avoiding and reducing harm is grounded in one of the most fundamental concerns in moral philosophy: 'the principle of non-maleficence'.[16] It states that one ought not to inflict harm on others. Paulo and Kirchmair frame it in relation to John Stuart Mill's harm principle, which holds that the 'only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others'.[17] While Mill's principle specifically concerns the justification of interference with individual liberty, it aligns with the broader ethical commitment to harm prevention expressed by the principle of non-maleficence.

What initially appears to be a convincing argument for the transition to self-driving cars can be challenged by reasonable objections. The argument heavily relies on the premise that self-driving cars will be safer. If this assumption holds, self-driving cars would be a key factor in achieving the European Union's road safety target of 'Vision Zero', which aims to eliminate all road fatalities by 2050. Yet even advocates

[13]Ibid 6.

[14]The Operational Design Domain (ODD) aims to define under which conditions an Automated Driving System (ADS) can operate safely, including factors like speed limits, road classifications, and environmental conditions. For further information, see L Mendiboure and others, 'Operational Design Domain for Automated Driving Systems: Taxonomy Definition and Application', *2023 IEEE Intelligent Vehicles Symposium (IV)* (2023).

[15]TA Dingus and others, 'Driver Crash Risk Factors and Prevalence Evaluation Using Naturalistic Driving Data' (2016) 113 Proceedings of the National Academy of Sciences 2636.

[16]N Paulo and L Kirchmair, *Moral Dilemmas Involving Self-Driving Cars: How to Regulate Them and Why Your Opinion Matters* (Routledge 2025) 6.

[17]JS Mill, *On Liberty* (first published 1859, Batoche Books 2001) 13.

of this ideal remain cautious about the impact such vehicles will have on traffic safety.[18] While they rightly follow Paulo and Kirchmair's argument that the significant prevalence of driver-related factors in crashes calls for technological support, this support need not take the form of full automation. Given the technical and practical challenges that continue to limit the deployment of automated vehicles, the authors conclude that a 'combination of the driver and the technology of a vehicle could under certain conditions be as safe as, or even safer than, the fully automated car'.[19]

Even if the risks introduced by automation itself are set aside, the transition to full road traffic automation will unfold over decades of mixed traffic, with automated vehicles sharing the road with human drivers.[20] Technico stresses that these transitional stages raise the most urgent question for his work: How can an automated system operate safely and fluently in traffic still shaped by human behaviour? As long as mixed traffic prevails, with human drivers interacting unpredictably with partially automated systems, the anticipated safety gains remain uncertain.

The same necessary caution must extend to the metric of accident reduction itself. A system optimised to minimise the overall number of crashes may still alter the distribution of harm in ways that matter ethically and politically. It could, for instance, simply shift exposure from vehicle occupants to vulnerable road users. Fewer accidents in total do not necessarily mean fairer outcomes. What appears as a safety improvement on aggregate may conceal new asymmetries of risk, depending on whose safety is being prioritised and whose is being compromised.

However, granting that the fair reduction of (physical) harm is certain, this narrow focus still overlooks other important dimensions of harm.[21] Harm might be caused in several other ways. As these cars might not be accessible or affordable to all, they could potentially discriminate against the poor or rural populations. Moreover, automation also threatens to displace millions of driving-related jobs, posing serious socio-economic challenges that must be addressed. Nevertheless, the impact of the fundamental principle of (physical) non-maleficence is clearly evident in the automotive industry, which holds safety as its undisputed highest value.

Given this clear commitment to minimising harm, Paulo and Kirchmair turn to the question of when it is permissible to redirect a self-driving car to that end. To address it, they explore established principles of moral philosophy, such as the Doctrine of Double Effect (DDE). It refers to a 'distinction between what a man foresees as a result of his voluntary action and what, in the strict sense, he intends'.[22] The basic notion is that it can be permissible to cause harm unintendedly or only as a foreseeable side effect of doing good, while it would be impermissible if this harm was used as a means to an end. Paulo and Kirchmair employ the doctrine to show that the principle of non-maleficence alone cannot determine what is morally permissible. Drawing on Foot's original examples, they contrast the case of a judge who considers executing an

[18] A Lie and others, 'Automated Vehicles: How Do They Relate to Vision Zero' in K Edvardsson Björnberg and others (eds), *The Vision Zero Handbook* (Springer 2023) 1057.
[19] Ibid 1069.
[20] On the challenges automation poses, see L Bainbridge, 'Ironies of Automation' (1983) 19 Automatica 775.
[21] For an overview of the various interpretations the harm principle has received, see, for instance, J Edwards, 'Harm Principles' (2014) 20 Legal Theory 253.
[22] P Foot, 'The Problem of Abortion and the Doctrine of the Double Effect' in *Virtues and Vices and Other Essays in Moral Philosophy* (Oxford University Press 2002) 19, 20.

innocent person to prevent riots with that of a trolley driver who can divert harm from five people to one. While the judge intends harm as a means to an end, rendering the act impermissible, the driver merely foresees it, making redirection morally acceptable.

Building on this, the authors further guide the reader through Foot's line of thinking as she concludes that 'the distinction between direct and oblique intention plays only a quite subsidiary role in determining what we say in these cases'.[23] Instead, she proposes a framework based on negative and positive duties, where duties not to harm (negative) are considered more morally stringent than duties to help (positive). Accordingly, the judge must refrain from killing the innocent person, as this would violate a negative duty. The driver, however, faces two conflicting negative duties, making it morally permissible to choose the lesser harm.

Gradually approaching the ethical challenges of autonomous vehicles, Paulo and Kirchmair now engage with Thomson's influential critique of Foot's argument. She introduces a case called '*Bystander at the Switch*'[24] that closely mirrors Foot's Driver case while introducing an important distinction: the agent is not the driver but a bystander capable of operating a switch to divert the trolley. According to Foot's framework, the bystander should refrain from pulling the switch, as the negative duty not to kill the one person outweighs the positive duty to save the five. Yet Thomson finds pulling the switch morally permissible.

To further explore this issue, Thomson presents a scenario that Paulo and Kirchmair refer to as the Footbridge case, where the agent has the option to push a large man off a footbridge and onto the tracks to stop the trolley. According to her, this action feels intuitively impermissible. This puzzling contrast between the permissibility in the Bystander case and the impermissibility in cases like Footbridge is what Thomson refers to as the 'The Trolley Problem'.[25] The reasons for this intuition are widely debated.[26] Paulo and Kirchmair underscore that '[t]he whole philosophical debate that evolved around Trolley cases assumes that these two judgments are correct'.[27] It focuses on the justification of these judgments, rather than questioning their legitimacy.

According to Paulo and Kirchmair, these philosophical discussions offer important insights for the ethics of autonomous vehicles, since they see an analogy between the discussed trolley cases and the scenarios involving self-driving cars. In their view, the Bystander case parallels situations in which an autonomous vehicle may swerve to save a larger group. To establish a connection to the Footbridge case, they devise their own scenario, which they call the 'Microcar case'.[28] An empty self-driving car's brakes fail while approaching a red light at an intersection, threatening five pedestrians. The only way to prevent their deaths is to redirect a small autonomous vehicle carrying a single passenger into its path, sacrificing him to stop the car and thereby save the pedestrians.

[23]Ibid 29.
[24]JJ Thomson, 'The Trolley Problem' (1985) 94 The Yale Law Journal 1395, 1397, emphasis in original.
[25]Ibid 1401.
[26]For a brief history of different interpretations of the problem and attempts to solve it, see FM Kamm, *The Trolley Problem Mysteries* (Oxford University Press 2016).
[27]N Paulo and L Kirchmair, *Moral Dilemmas Involving Self-Driving Cars: How to Regulate Them and Why Your Opinion Matters* (Routledge 2025) 36.
[28]Ibid 35.

Paulo and Kirchmair conclude that, in a dilemma situation, redirecting the vehicle to avoid greater harm is considered permissible, while actively sacrificing an individual – whether standing on a bridge or seated in a microcar – is not. Although they acknowledge that none of the explanations of these judgements 'convinced many philosophers'[29], they do not unpack the ramifications of this unresolved debate. One can either attempt to find a widely acceptable explanation for this judgment – treading a path where many great minds have previously faltered – or abandon the insufficiently grounded intuition altogether. Practically, this abandonment would mean no longer treating vehicle redirection as a viable option. Thomson herself took this exact step of abandonment over 30 years after her famous paper.[30]

Technico raises two additional concerns about Paulo and Kirchmair's discussion. First, the authors' use of idealised scenarios and optimistic claims about technological progress risks distorting public understanding of what autonomous vehicles can actually achieve. Second, recommendations derived from such thought experiments are ill-suited to the complexities of technical realities.

As for the first concern, the authors acknowledge that the Microcar case is 'somewhat far-fetched'[31], yet they do not deem it to be entirely unrealistic. For their book, Paulo and Kirchmair assume that self-driving vehicles are at the highest level of automation and therefore operate everywhere, without human intervention.[32] When they talk about tremendous improvements in the technology over the next few years, they are referring to what has been 'promised by the automotive industry'.[33] These claims about technological advancement are infused with marketing strategies aimed at attracting potential customers and investors. Reports celebrating milestones in autonomous driving technology often omit a key detail: the Operational Design Domain (ODD). It is a false assumption that a car capable of driving autonomously in one specific environment will function equally well in other – or even all – domains. This is what I call the 'ODD fallacy'.

This fallacy has significant practical implications. Creating the impression that advanced driver assistance systems can handle scenarios beyond their actual capabilities entails significant risks. A terrible accident recently motivated the Chinese government to restrict questionable marketing strategies, 'direct[ing] carmakers to stop using the terms "smart driving" and "autonomous driving" in advertisements for driver assistance systems'.[34]

While Paulo and Kirchmair rightfully point out that the interconnectivity of autonomous vehicles is one of the 'main tools to make them drive more safely than human drivers'[35], they significantly overestimate what the technology can actually achieve.

[29]Ibid.
[30]See JJ Thomson, 'Turning the Trolley' (2008) 36 Philosophy & Public Affairs 359.
[31]N Paulo and L Kirchmair, *Moral Dilemmas Involving Self-Driving Cars: How to Regulate Them and Why Your Opinion Matters* (Routledge 2025) 35.
[32]For an explanation of all levels of automation defined by the Society of Automotive Engineers (SAE), see On-Road Automated Driving (ORAD) Committee, 'Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles' (SAE International 2021) <https://www.sae.org/content/j3016_202104> accessed 10 July 2025.
[33]N Paulo and L Kirchmair, *Moral Dilemmas Involving Self-Driving Cars: How to Regulate Them and Why Your Opinion Matters* (Routledge 2025) 14.
[34]P Shukla, 'China Bans "smart Driving" Ads after Fatal Crash Involving Xiaomi EV' (17 April 2025) <https://mybs.in/2emGjtU> accessed 10 July 2025.
[35]N Paulo and L Kirchmair, *Moral Dilemmas Involving Self-Driving Cars: How to Regulate Them and Why Your Opinion Matters* (Routledge 2025) 35.

From their perspective, 'each car will "know" the traffic situation everywhere along its route, far beyond what can be "seen" with its sensors, let alone with human eyes'.[36] These cars will be able to 'anticipate dangerous situations much better than human drivers can'.[37] The authors take it one step further by stating that self-driving vehicles will 'basically have all of the information about all of the other cars on the road, both past and present'[38], including where and even why accidents have happened. This idea of a machine learning algorithm improving itself with its experiences in traffic omits the issues in terms of explainability related to such an approach. In order to meet regulatory and ethical requirements, these algorithms would have to be tested at great length relying on 'a precise definition of what constitutes ethically acceptable behavior'.[39] Such a definition has yet to be found. Given these often overlooked challenges, Technico formulates the following principle: *Public engagement with autonomous vehicle ethics must be grounded in a realistic assessment of technical capabilities.*

Technico's second concern turns to a related but distinct issue: the effectiveness of the authors' purportedly practical recommendations. While there might be relevant parallels between the moral dilemmas presented by trolley cases and those involving self-driving cars, the practical relevance of their discussion diminishes when confronted with the complexities of real-world application. Paulo and Kirchmair are aware that self-driving cars will not deal with outcomes that are 'fixed and known with certainty'.[40] Yet, they defend the value of these thought experiments. By using a phrase from Brown, the authors propose to think of it as a 'laboratory of the mind'.[41] They say, 'As a moral issue, the trolley problem is interesting irrespective of all of the other factors that potentially determine what happens in real-life accident situations.'[42] Paulo and Kirchmair thereby handle the objections by attacking their relevance. According to them, realistic scenarios would bear the risk of losing sight of the precise moral issue. They add, 'Trust us, if it were easier to think through difficult moral problems in complex real-life situations, philosophers would be thrilled to do just that.'[43] Within the ivory tower of ethical theorizing, this argument would not pose any problems. However, tension arises when a book that aims to provide practical recommendations marginalises practical concerns.

Even so, the authors briefly touch on the subject of probabilities. For them, '[t]he moral issue is not what the chances are that the object really is a human being or that the crash will kill a person'.[44] In consequence, Paulo and Kirchmair make clear that situations in which it is highly unlikely that someone will die is not a dilemma they are interested in. By deliberately narrowing their focus, they overlook at least two issues that play a crucial role in practice and must be addressed in a meaningful legislation: (i) What

[36]Ibid.
[37]Ibid.
[38]Ibid 9.
[39]LR Sütfeld, J Bronson and L Kirchmair, 'Automated Vehicle Regulation Needs to Speak to Code, Not to Humans: Keeping Safety and Ethics in the Public Domain' (2025) 38 Philosophy & Technology 15, 15.
[40]N Paulo and L Kirchmair, *Moral Dilemmas Involving Self-Driving Cars: How to Regulate Them and Why Your Opinion Matters* (Routledge 2025) 37.
[41]JR Brown, *The Laboratory of the Mind: Thought Experiments in the Natural Sciences* (Routledge 2011).
[42]N Paulo and L Kirchmair, *Moral Dilemmas Involving Self-Driving Cars: How to Regulate Them and Why Your Opinion Matters* (Routledge 2025) 38.
[43]Ibid.
[44]Ibid 39.

moral significance do probabilities have in a necessary ex ante assessment? (ii) How can injuries of varying severity be weighed against each other? Technico therefore sets out a second principle: *Practical recommendations must account for real-world complexity*.

Despite these difficulties, the trolley cases effectively illustrate fundamental moral principles relevant to self-driving vehicles. Given the extensive debate surrounding such cases over the decades, it is understandable why Paulo and Kirchmair refer to them as 'philosophical gold'.[45] This view, however, is not universally shared.[46] Eventually, the social sciences also recognised the potential of these cases, posing the question: How would laypeople decide in these kinds of situations? Paulo and Kirchmair seek to draw on these additional insights.

3. Guardrails for public opinion

A large-scale study called 'The Moral Machine experiment' elicited 40 million responses to moral dilemmas involving self-driving cars via an online platform, in ten languages and from millions of people in 233 countries and territories.[47] Its findings have sparked a major debate on the significance of laypeople's moral preferences for the programming of self-driving vehicles. While in the philosophical discussion Paulo and Kirchmair focused on justifying intuitions with moral principles, their analysis of the empirical research shifts the focus to assessing the reliability of these intuitions, questioning whether they can serve as a stable basis for moral judgment or policy.

First, Paulo and Kirchmair show that the intuitions regarding the Bystander and Footbridge cases are reflected in empirical data. They cite a study in which approximately 90% consider flipping the switch permissible, while only approximately 10% view pushing the man off the bridge as morally acceptable.[48] As a result, programming a self-driving vehicle to swerve in order to save the many appears supported by both philosophical debate and empirical evidence. But one major question still persists: Where to swerve to?

Notably, while the original philosophical discussions of these dilemmas did not focus on *who* was saved, the empirical experiments incorporated personal characteristics into the evaluation. Paulo and Kirchmair focus their discussion on three of these factors that significantly influenced the study participants' decisions on how to steer the vehicle: social status, age and gender.

The authors point out that the notion of preferential treatment is not without precedent. In an interesting historical detour, they show how triage decisions during events such as the French Revolutionary Wars, the Titanic disaster, and the COVID-19 pandemic were based on personal characteristics. Faced with this tradition, Paulo and Kirchmair question whether such considerations belong in the regulatory context of self-driving cars: 'Are these the wrong questions to ask?'[49] They find their doubts echoed in a remark by Bonnefon, one of the psychologists involved in the Moral

[45]Ibid 34.
[46]See, for instance, FM Kamm, 'The Use and Abuse of the Trolley Problem: Self-Driving Cars, Medical Treatments, and the Distribution of Harm' in SM Liao (ed), *Ethics of Artificial Intelligence* (Oxford University Press 2020) 79.
[47]E Awad and others, 'The Moral Machine Experiment' (2018) 563 Nature 59.
[48]M Hauser and others, 'A Dissociation Between Moral Judgments and Justifications' (2007) 22(1) Mind & Language 1.
[49]N Paulo and L Kirchmair, *Moral Dilemmas Involving Self-Driving Cars: How to Regulate Them and Why Your Opinion Matters* (Routledge 2025) 28.

Machine experiment. He explains that the category of social status was included to highlight the limitations of using such experiments to inform public policy.[50]

The Moral Machine and the different historical cases of preferential treatment have one thing in common: the agent is forced to decide. To illustrate the significance of this aspect, the authors draw the reader's attention to a much smaller study with fewer than 1000 participants.[51] This study allowed for an option of equal treatment, meaning a refusal to favour one side over the other. When this choice was available, the proportion of people who preferred to save those of high social status dropped from 79.7% to just 12.7%. Similar results were observed in the case of other personal characteristics. But Paulo and Kirchmair stress that there is one decisive preference that still remains: Saving the many is 'one of the most dominant moral preferences to be found worldwide'.[52] In their epilogue, they write: 'If we had to pin down the most important finding of this book, perhaps it would be that most people have the moral preference that it is right to choose to protect the larger group over the smaller group.'[53] They refer to Bigman and Gray in support of this claim, noting that even if people are given the option to treat people equally 'only 17.9% preferred this equality option; a majority (81.6%) still preferred to save as many lives as possible'.[54]

However, a closer look at the study and its supplementary material suggests a possible misreading: in Study 2 – the relevant one given the introduction of the equality option – the actual figures appear to be reversed. According to the data presented on page 6 of the Supplementary Information, 81.6% of participants preferred equal treatment, while only 17.9% maintained a preference for saving more lives. If this interpretation holds, it undermines a central empirical pillar of the authors' argument and calls for clarification or correction.

Regardless of their results, the authors conclude that there is a great need for 'a critical evaluation of the methodological approaches'.[55] Beyond established scientific standards like adequate sample sizes, replicability, and representativeness, the effect of including an equality option highlights the importance of thoughtful study design. Such design choices play a crucial role in addressing potential biases. Paulo and Kirchmair effectively illustrate these various challenges, drawing on the influential work of Tversky and Kahneman[56], among others.

Although a detailed account of this discussion cannot be provided here, Technico points to one important aspect missing from Paulo and Kirchmair's analysis. Preferential treatment in autonomous vehicle technology can arise for various technical reasons. Some of these effects can only be mitigated, not fully overcome – even when public preferences demand it. The authors assume that self-driving vehicles offer the opportunity to 'program the outcome of such accidents in a thoughtful, well-informed, and deliberate

[50]J-F Bonnefon, *The Car That Knew Too Much: Can a Machine Be Moral?* (The MIT Press 2021).
[51]YE Bigman and K Gray, 'Life and Death Decisions of Autonomous Vehicles' (2020) 579 Nature E1.
[52]N Paulo and L Kirchmair, *Moral Dilemmas Involving Self-Driving Cars: How to Regulate Them and Why Your Opinion Matters* (Routledge 2025) 36.
[53]Ibid 85.
[54]Ibid 71.
[55]Ibid 55.
[56]A Tversky and D Kahneman, 'Judgment under Uncertainty: Heuristics and Biases: Biases in Judgments Reveal Some Heuristics of Thinking under Uncertainty.' (1974) 185 Science 1124.

way'.[57] This expectation extends beyond what is technically feasible. Even if the '[c]omputational capacity far outstrips human capacity when reacting in an accident'[58], other factors must be considered as well. The trajectory planning is only one aspect of the 'standard system architecture for automated driving systems to date'.[59]

Before turning to the other aspects, it is important to clarify the ethical concern at stake. Earlier discussions focused on preferential treatment, meaning decisions that consciously favour one group over another. In technical systems, the concern is better described as unequal treatment, which may or may not constitute discrimination. In this context, I use 'discrimination' to refer to wrongful unequal treatment which can arise either intentionally or unintentionally. If it is rooted in intentional bias, meaning the systematic focus on certain groups, or negligence in addressing foreseeable risks, the need for correction is evident. In contrast, unintended discriminatory effects require careful analysis. Clarifying their cause is essential for determining appropriate mitigation measures and regulatory thresholds. This demands distinguishing between technical constraints and bias, which stems from how systems are designed, trained, or evaluated. These differences become apparent when examining the technical implementation in more detail.

From a technical perspective, situations that seem to discriminate against a specific group of people can be caused on at least three different levels: Sensors, data, and algorithms. At the sensor level, physical constraints may disadvantage certain individuals – such as children – since smaller 'objects' are harder to detect. Although these technical causes do not justify the resulting disadvantages, they constrain the extent to which such effects can be mitigated. At the data level, bias can arise when training datasets underrepresent specific groups. To improve their detection, data must be labelled to distinguish specific categories of people, allowing models to prioritise their recognition. At the algorithmic level, comparable challenges emerge, as the outcome hinges on how models are optimised and assessed. Assigning greater weight to rare cases can mitigate bias, but without disaggregated assessment, for example by skin tone or size, unequal treatment may remain hidden.

Since these challenges will inevitably persist, the question then is: What level of unequal treatment should be deemed permissible? Current systems cannot reliably detect personal characteristics such as age, social status, or gender, regardless of their relevance in empirical studies. Nor can they always accurately determine the number of individuals involved.[60] Given these considerations, Technico articulates a third principle: *Policymakers must determine whether, and to what extent, unequal treatment is acceptable in light of technical limitations.*

4. Driving toward equilibrium

A central contribution of Paulo and Kirchmair's book is its compelling argument that laypeople's moral intuitions should inform the regulation of self-driving cars. If diligently put into practice, this is a welcome claim. At a time when many policy and industry

[57]N Paulo and L Kirchmair, *Moral Dilemmas Involving Self-Driving Cars: How to Regulate Them and Why Your Opinion Matters* (Routledge 2025) 9.
[58]Ibid 51.
[59]LR Sütfeld, J Bronson and L Kirchmair, 'Automated Vehicle Regulation Needs to Speak to Code, Not to Humans: Keeping Safety and Ethics in the Public Domain' (2025) 38 Philosophy & Technology 15, 4.
[60]For an insight into the problems of pedestrian detection with occlusion, see, for instance, C Ning and others, 'Survey of Pedestrian Detection with Occlusion' (2021) 7 Complex & Intelligent Systems 577.

debates around AI ethics are dominated by technocratic expertise or market-driven decision-making, the authors insist that questions of justice and risk allocation in traffic must be resolved through 'broad public debate'.[61]

To support their claim, the authors draw on a selection of philosophers who consider public preferences to be of 'prime importance for questions of justice'.[62] They side with David Miller when he stresses that normative theorizing ultimately aims at providing practical guidance and must therefore be 'accessible to the relevant agents'.[63] However, according to Paulo and Kirchmair, in the contested relationship between academic normative theorizing and practical application, the emphasis typically lies on the theoretical side. To advocate for a revision of these tendencies, they draw on Swift and White, adapting their description of the political theorist as an aid to democratic deliberation to the role of the philosopher.[64] In Paulo and Kirchmair's view, the philosopher should contribute by 'offering help in formulating principles and policies, indicating their implications, offering criticism, more arguments, and justifications'.[65]

Despite these efforts, allocating greater weight to empirical findings carries the inherent risk of conflating public moral preferences with normative justification. As scholars in experimental philosophy have argued, empirical data on moral intuitions may reveal how people in fact judge, but not what they are justified in judging. Attempts to infer normative conclusions from such data risk committing an is-ought fallacy. Yet, as Mortensen and Nagel demonstrate, experimental and traditional approaches need not stand in opposition.[66] Their work exemplifies how descriptive studies can complement, rather than replace, normative analysis.

To avoid the impression that their project seeks the replacement of normative reasoning with empirical observation, Paulo and Kirchmair begin their book by outlining the normative landscape of self-driving cars. Only then do they introduce empirical findings – not to replace theory with data, but to highlight the tensions between the two and ultimately integrate them through a structured method.

In their search for such a method, the authors recognise great potential in an approach called 'Collective Reflective Equilibrium in Practice (CREP)'.[67] Its aim is to show how 'data about public preferences may be used to inform policy around the use of controversial novel technologies'.[68] Inspired by a decision procedure for ethics outlined by John Rawls[69], Savulescu, Gyngell and Kahane present a way of combining moral theories and public moral preferences. The authors of CREP describe the original notion of a

[61]N Paulo and L Kirchmair, *Moral Dilemmas Involving Self-Driving Cars: How to Regulate Them and Why Your Opinion Matters* (Routledge 2025) 13.
[62]Ibid 55.
[63]D Miller, 'Needs-Based Justice: Theory and Evidence' in AM Bauer and M Meyerhuber (eds), *Empirical Research and Normative Theory: Transdisciplinary Perspectives on Two Methodical Traditions Between Separation and Interdependence* (De Gruyter 2020) 273, 274.
[64]See A Swift and S White, 'Political Theory, Social Science, and Real Politics' in D Leopold and M Stears (eds), *Political Theory: Methods and Approaches* (Oxford University 2008) 49.
[65]N Paulo and L Kirchmair, *Moral Dilemmas Involving Self-Driving Cars: How to Regulate Them and Why Your Opinion Matters* (Routledge 2025) 56.
[66]K Mortensen and J Nagel, 'Armchair-Friendly Experimental Philosophy' in J Sytsma and W Buckwalter (eds), *A Companion to Experimental Philosophy* (Wiley 2016).
[67]J Savulescu, C Gyngell and G Kahane, 'Collective Reflective Equilibrium in Practice (CREP) and Controversial Novel Technologies' (2021) 35 Bioethics 652.
[68]Ibid.
[69]See J Rawls, 'Outline of a Decision Procedure for Ethics' (1951) 60 The Philosophical Review 177.

'reflective equilibrium'[70] as being focused on '[e]thical justification for judgments about specific cases and revision of general principles and theories'.[71] However, they extend its application to the justification of policy within a democratic system. In their effort of 'integrating ethical theory and data about the spread of public intuitions while minimizing the risks associated with each'[72], the authors must carefully consider what to include in each domain.

First, they need to decide which theories and principles to incorporate. As Savulescu and his co-authors explain, there are two approaches: draw on widely shared public values or on 'ethical theories that, after decades of critical reflection, are seen as serious candidates within moral philosophy'.[73] While they favour the latter, they emphasise the importance of balancing both approaches to avoid relying exclusively on either abstract ideals that lack public support or widely held views that may be 'pernicious'.[74]

Second, they must determine what constitutes sufficiently 'laundered preferences'[75], meaning what judgements should be allowed entry to the process. As previously noted, addressing bias requires not only standard methodological safeguards but also deliberate design features – such as offering an equality option – to ensure that empirical studies reflect ethically meaningful preferences.

Technico insists that this laundering must also extend to the study's setting. It should assess the real-world applicability of its scenarios in light of technical constraints – such as the difficulty of detecting personal characteristics or estimating probabilities in real time. Empirical studies that aim to inform policy should ensure that their scenarios reflect what is technologically relevant and realistically achievable. This leads Technico to formulate his fourth principle: *Policymakers must base their decisions on studies grounded in the technical realities and practical complexity of the systems they aim to regulate*.

Given the global use of autonomous vehicles, it is important to recognise that ethical perspectives may vary by cultural context. Paulo and Kirchmair draw on the work of Joseph Henrich to present this challenge. Henrich uses the acronym 'WEIRD' to describe people from societies that are 'Western, Educated, Industrialized, Rich, and Democratic'.[76] While the Moral Machine experiment grouped results into three clusters – Western (North America, many European countries, and some Commonwealth nations), Eastern (many Far East countries), and Southern (Latin America and French-influenced nations) – no distinct African cluster emerged due to limited participation from these countries.

Paulo and Kirchmair do not address a fundamental question for practical implementation that arises here. As Technico points out, if moral judgments vary across cultural contexts, how should self-driving vehicles be programmed to respect these differences? Does this require developing country-specific versions? The dominance of Western

[70]See J Rawls, *A Theory of Justice* (Harvard University Press 1971).
[71]J Savulescu, C Gyngell and G Kahane, 'Collective Reflective Equilibrium in Practice (CREP) and Controversial Novel Technologies' (2021) 35 Bioethics 652, 657.
[72]Ibid 656.
[73]Ibid 659.
[74]Ibid.
[75]Ibid 657.
[76]J Henrich, *The WEIRDest People in the World: How the West Became Psychologically Peculiar and Particularly Prosperous* (Farrar, Straus and Giroux 2020) iii.

cultures in these studies makes this question particularly salient. Technico therefore advances a fifth principle: *Policymakers must account for the diverse cultures in which autonomous vehicles operate*.

After establishing what is to be included in the method, Paulo and Kirchmair describe the process of reaching a conclusion as relatively straightforward. According to them, it can be understood as follows: 'support for a policy is greater the more it reflects public preferences and the more it coheres with moral theories'.[77] However, their account leaves the process of balancing these two elements largely unspecified. Unlike traditional reflective equilibrium, where moral theories and judgments are mutually adjusted, their approach risks appearing additive rather than integrative. But how to deal with cases that do not show clear rejection or clear approval? Here, one of the most important contributions from Paulo and Kirchmair's book comes into play.

The authors expand the approach by introducing a third element. They 'propose extending the CREP method to include human rights as a tie-breaker'.[78] To support this proposal, Paulo and Kirchmair provide several reasons. First, they point out that human rights reflect a form of a 'long-established international agreement on fundamental values' reached through 'long and arduous social processes'.[79] While moral preferences do not have to be in alignment with the law, the authors argue that 'we should be concerned when they violate human rights'.[80] They assume that their elevated role in the method should arouse little opposition, as human rights are an expression of the most fundamental values.

Bearing in mind Henrich's work on the WEIRD perspective, this assumption does not necessarily hold. Paulo and Kirchmair briefly acknowledge that the acceptance and implementation of human rights may not be as universal in practice as sometimes assumed. As a mitigation, they suggest that whenever there is a 'focus on public preferences found only in studies with a regional focus, regional human rights should serve as a reference point'.[81] However, relying on human rights as a shared foundation faces practical challenges. While most countries in the world have ratified the major human rights treaties, this acceptance does not necessarily guarantee appropriate interpretation, let alone implementation.[82]

Nevertheless, Paulo and Kirchmair's proposal is compelling due to its added gatekeeper function which further mitigates the risk of conflating empirical findings on public moral preferences with normative justification. While the CREP method already permits the inclusion of human rights – referring broadly to 'ethical frameworks'[83], including theories, principles, concepts, and professional guidelines – Paulo and Kirchmair's contribution lies in emphasising human rights as a central component

[77]N Paulo and L Kirchmair, *Moral Dilemmas Involving Self-Driving Cars: How to Regulate Them and Why Your Opinion Matters* (Routledge 2025) 64.
[78]Ibid 66.
[79]Ibid 65.
[80]Ibid.
[81]Ibid.
[82]For an analysis of the power-related aspects of human rights, see P Gilabert, 'Reflections on Human Rights and Power' in A Etinson (ed), Human Rights: Moral or Political? (Oxford Academic 2018).
[83]J Savulescu, C Gyngell and G Kahane, 'Collective Reflective Equilibrium in Practice (CREP) and Controversial Novel Technologies' (2021) 35 Bioethics 652, 661.

that merits distinct acknowledgment. The authors illustrate its value 'in the event of unclear outcomes produced by CREP'.[84]

Take, for instance, the category of age. The Moral Machine experiment identified favouring the young as the third strongest global preference, but in Bigman and Gray's study 61.1% preferred treating individuals of all ages equally. These results are far from reflecting universal agreement. Paulo and Kirchmair find that established moral theories paint a similar picture. While Kantian deontology 'clearly prohibits any distinction based on personal features', a utilitarian perspective 'would prioritize younger people, because, on average, younger lives lead to more utility in the future'.[85] The authors conclude that '[t]he moral verdict is thus unclear'.[86] Finally, human rights offer guidance. Paulo and Kirchmair draw attention to the 'human right to equality, and particularly the obligation of non-discrimination'.[87] They note that this principle includes a 'ban on age discrimination in most jurisdictions'.[88] Human rights thus tilt the balance towards equality and therefore not taking age into account.

This example prompts further inquiry into the conceptual distinction between moral theories and human rights, since both may prohibit discrimination based on personal characteristics and can thereby lead to the same normative conclusion. While the appeal to human rights may strengthen the legal authority of this position, it does not clearly add a new moral dimension. The appeal to human rights may not constitute an independent source of justification but rather reinforces conclusions already grounded in prior ethical reasoning. An explicit discussion of this relationship would have strengthened the conceptual clarity of the authors' proposal.

Moreover, it remains unclear how effective this approach would be in resolving cases where multiple policies cohere with moral theories and public preferences without violating human rights. The tie-breaker may rule out certain unacceptable options, but it does not necessarily point to a single correct policy.

This interplay of normative principles and public preferences culminates in the pressing question of how such considerations can be embedded within concrete regulatory frameworks.

5. Braking the trolley

In their concluding chapter, Paulo and Kirchmair confront the challenge of embedding their insights into regulatory frameworks, offering practical recommendations for policymakers. While each of these proposals warrants further discussion, I limit my analysis to a specific aspect of autonomous vehicle regulation that the authors have largely overlooked.

Much of the authors' policy advice builds naturally on their earlier conclusions, such as the urgency of regulation and the need to minimise harm irrespective of social status, age, or gender. However, the authors also introduce claims that are – although crucial –

[84]N Paulo and L Kirchmair, *Moral Dilemmas Involving Self-Driving Cars: How to Regulate Them and Why Your Opinion Matters* (Routledge 2025) 66.
[85]Ibid 70.
[86]Ibid.
[87]Ibid.
[88]Ibid 71.

only inadequately addressed. Paulo and Kirchmair subsume them under one central principle: 'Regulation needs to talk to code'.[89]

In the corresponding sub-chapter, they present five demands. First, they state that '[r]egulations for machines must be designed differently than regulations for humans'[90], since machines require precise instructions. Despite this requirement, the authors argue that, second, a balance must be struck here: While regulation should constrain corporate self-interest through clearly specified requirements, it must also allow for sufficient leeway to develop optimal technical solutions. Third, the authors ask for the regulation to 'take place, time, action, and responsibility into account as well as various options and probabilities'.[91] With their fourth proposition, 'probabilities matter'[92], they emphasise this last part once again. At this point, it is worth remembering how little relevance the authors assigned to probabilities in their defence of the trolley cases. The fifth and final demand can be seen as a consequence of the previous ones: In the face of all these enormous challenges, it is necessary to ensure that guidance is provided to balance all these aspects. Unfortunately, Paulo and Kirchmair do not discuss these aspects in greater detail. While their agenda seems reasonable, it marks a notable departure from the authors' previous line of thinking. By aiming at *practical* advice, they leave the theoretical domain that they defended so decisively earlier. However, their discussion remains too cursory to address the practical challenges convincingly.

The authors' reference to a recent publication by Sütfeld, Bronson, and Kirchmair may serve to compensate for these shortcomings, as it offers a valuable account of the 'tension between the current state of the law and the realities of trajectory-related decision-making in AVs'.[93] Whereas the paper limits its analysis to German and European law on autonomous vehicles, Technico must contend with a broader regulatory landscape. Since Paulo and Kirchmair assume that self-driving vehicles operate at the highest level of automation, they overlook a more immediate concern: the challenge of obtaining certification for those levels defined by the Society of Automotive Engineers (SAE) that are realistically attainable today. To this end, Technico draws attention to the regulatory framework set out by the United Nations Economic Commission for Europe (UNECE), particularly Regulation R171 on Driver Control Assistance Systems (Level 2) and Regulation R157 on Automated Lane Keeping Systems (Level 3).[94] There is good reason to expect that the regulatory demands at lower levels of automation will shape the course of autonomous vehicle development. In this regard, R157 contains a particularly noteworthy requirement concerning how a vehicle should respond in the event of an impending accident – closely resembling the scenarios discussed by Paulo and Kirchmair.

It allows for evading under three conditions: (i) the manoeuvre is at least as safe to the vehicle occupants and other road users as avoiding the imminent collision risk by braking; (ii) the manoeuvre does not cause a collision with another vehicle or road user in the predicted path of the vehicle; (iii) the manoeuvre does not force another

[89]Ibid 83.
[90]Ibid.
[91]Ibid.
[92]Ibid.
[93]LR Sütfeld, J Bronson and L Kirchmair, 'Automated Vehicle Regulation Needs to Speak to Code, Not to Humans: Keeping Safety and Ethics in the Public Domain' (2025) 38 Philosophy & Technology 15, 3.
[94]UNECE vehicle regulations are binding within the EU legal framework through their incorporation into Regulation (EU) 2018/858.

vehicle in the target lane to unmanageably decelerate.[95] These conditions impose demanding requirements on a vehicle to justify initiating evasive action.

At first glance, these provisions may seem to contrast with the German Act on Autonomous Driving of 2021, cited by the authors.[96] The latter demands an accident-avoidance system that fulfils three requirements: First, it is designed to avoid and reduce harm. Second, in the event of unavoidable alternative harm to different legal interests, the importance of legal interests needs to be considered, with the protection of human life as the highest priority. Third, in the case of unavoidable alternative harm to human life, there is no further weighting on the basis of personal characteristics.

While the Act appears more permissive in allowing value-based trade-offs, both frameworks ultimately converge on a common objective: the minimisation of harm. Whereas R157 explicitly outlines the conditions under which evasive action may be taken, the Act implicitly calls for a comparable evaluation of potential harm, as its first requirement is to avoid and reduce it. To meet this requirement, evasive action can only be considered if its outcome, specifically the extent to which harm is avoided or reduced, is sufficiently foreseeable. This demand is grounded in both regulatory frameworks.

While one might object that this comparison is inappropriate – since R157 applies to Level 3 systems, whereas the Act on Autonomous Driving addresses Level 4 – the distinction becomes largely theoretical in emergency situations. Formally, Level 3 requires the human passenger to remain attentive and ready to resume control upon a take-over request. In practice, however, issuing such a request mere moments before a collision serves no meaningful purpose. Ultimately, at both levels, the vehicle itself must make the critical decision.

Justifying such a decision requires addressing several complex questions. For instance, how much harm may justifiably be caused to innocent parties – such as those in an adjacent lane – to prevent the probable loss of life within the vehicle's own lane? Trolley cases assume that redirecting the vehicle is a viable option, but the increased uncertainty calls this into question. For evasive action to be legitimate, the system must reliably estimate the additional risk of redirecting the vehicle and determine whether that risk falls within acceptable limits. Executing such a manoeuvre introduces significant challenges, as other traffic participants – both detected and undetected – may behave unpredictably. This unpredictability generates a degree of chaos beyond the system's control.

These challenges become particularly apparent in the case of leaving the road. Technico stresses that this would involve varying surface conditions, reduced vehicle control, the risk of rollover, collisions with obstacles, and other hazards. The system must be able to assess whether the manoeuvre remains controllable and whether leaving the road is likely to result in other accidents – and how severe their consequences might be.[97] Until such uncertainties can be adequately managed, the appropriate response to these moral dilemmas appears to be braking.

[95] See Inland Transport Committee of the Economic Commission for Europe, 'Uniform provisions concerning the approval of vehicles with regard to Automated Lane Keeping Systems' (2022), ECE/TRANS/WP.29/2022/59/Rev.1 par. 5.3.5.

[96] There is no official translation of the German Road Traffic Act (StVG). Paulo and Kirchmair are mainly referring to § 1e (2) No. 2, using an unofficial translation, which I have adopted here.

[97] Controllability and severity are two of the three crucial factors in evaluating the Automotive Safety Integrity Level (ASIL). The third factor is exposure, meaning the probability of the situation occurring. See International Organization for Standardization, *ISO 26262: Road Vehicles — Functional Safety* (2018).

Perhaps surprisingly, this approach aligns with the set of rules formulated by Paulo and Kirchmair. They ask for risk reduction, prioritising humans, disregarding personal characteristics, saving the greatest number, respecting traffic rules, and limiting active interference and deliberate killing as much as possible. While the authors present the preference for saving the greater number as one of their most important findings, they frame risk reduction as the 'the overarching principle'.[98] This analysis makes clear that braking – rather than evading – is the most effective means to adhere to this principle, as it minimises uncertainty and avoids introducing further uncontrollable risks.

This line of reasoning is further supported by the authors' rule that traffic rules should be respected. If trolley-like dilemmas were to occur in road traffic, they would most likely be the result of unlawful behaviour. Resolving these situations by weighing the lives of law-abiding individuals against those whose unlawful actions caused the danger appears ethically questionable. This moral preference was also found in the Moral Machine experiment, where it was as prominent as the preference for saving the young. Despite its prominence, Paulo and Kirchmair do not discuss it. This tendency is further supported by Bigman and Gray's study, in which 53.1% of participants still preferred to spare those who complied with the rules, even when an equality option was available.

Finally, in advocating to reduce interference and avoid deliberate killing as much as possible, Paulo and Kirchmair describe inaction – meaning to refrain from evasive action – as a conscious decision to let 'fate'[99] take its course. Perhaps, after all, the simplest solution is also the most prudent one. After extensively discussing and turning the trolley, it may be time to focus on braking it. This conclusion cannot only be aligned with the authors' own recommendations, but is also supported by the results of empirical studies and emerging legislative frameworks. Technico thereby formulates his final principle: *In crash scenarios, braking – rather than evasive action – must be the default response*.

6. Conclusion

Moral Dilemmas Involving Self-Driving Cars is a valuable and timely contribution to the debate on how to ethically and legally govern autonomous vehicle technology. Paulo and Kirchmair deserve particular praise for their interdisciplinary and inclusive approach, successfully bridging moral philosophy, legal theory, and empirical research. They lay out why and how public opinion should inform policy-making with both care and clarity. By introducing human rights as a tie-breaker, they enhance the promising CREP method. While their book serves as a good starting point for further engaging with the debates surrounding autonomous vehicles, its recommendations would benefit from greater attention to real-world applicability. As shown through the technical insights from the perspective of Technico, the omission of certain technical constraints leads to serious vulnerabilities in their argument.

[98]N Paulo and L Kirchmair, *Moral Dilemmas Involving Self-Driving Cars: How to Regulate Them and Why Your Opinion Matters* (Routledge 2025) 84.
[99]Ibid.

All in all, Paulo and Kirchmair's work offers a compelling foundation for future inquiry. Building on this, incorporating the perspective of an engineer holds promise that ethical frameworks for autonomous vehicles translate into effective and feasible policy. To this end, Technico's six principles guide the translation of ethical theory into technical and regulatory practice:

(1) Public engagement with autonomous vehicle ethics must be grounded in a realistic assessment of technical capabilities.
(2) Practical recommendations must account for real-world complexity.
(3) Policymakers must determine whether, and to what extent, unequal treatment is acceptable in light of technical limitations.
(4) Policymakers must base their decisions on studies grounded in the technical realities and practical complexity of the systems they aim to regulate.
(5) Policymakers must account for the diverse cultures in which autonomous vehicles operate.
(6) In crash scenarios, braking – rather than evasive action – must be the default response.

Future research must continue this dialogue to shape meaningful, ethically sound, and publicly legitimate policy on self-driving vehicles. In this sense, Technico does more than provide a pragmatic footnote. He complements the dialogue initiated by Principia and Skeptico by representing the voice of technical feasibility. His presence ensures that ethical and legal reasoning remains responsive to technical realities, so that principles are not only persuasive in theory but also viable in practice.

Acknowledgements

I would like to thank my colleagues at the Institute for Ethics in Technology at Hamburg University of Technology (TUHH) for their insightful discussions and support. I am also grateful to the anonymous reviewer for valuable comments that helped to improve this essay.

Disclosure statement

The author is employed by Mercedes-Benz Group AG in the field of automated driving ethics. This article was written in the author's capacity as a PhD candidate at the Hamburg University of Technology (TUHH). The views expressed are solely those of the author and do not necessarily reflect the views of Mercedes-Benz Group AG.

ORCID

Jan Hölzer http://orcid.org/0009-0003-2551-3128