Title Page

Title: Impartiality Preferences in Sacrificial Moral Dilemmas Involving Autonomous Vehicles

Word count: 3,998 (excluding bibliography)

Authors:

1. Norbert Paulo, School of Transformation and Sustainability, Catholic University of Eichstätt-Ingolstadt & Department of Social Sciences and Public Affairs, University of the Bundeswehr Munich, Germany, ORCID: 0000-0001-9644-6194; Norbert.paulo@gmail.com
2. Lando Kirchmair, Department of Social Sciences and Public Affairs, University of the Bundeswehr Munich, Germany, ORCID: 0000-0002-1924-1352; lando.kirchmair@unibw.de
3. Yochanan E. Bigman, The Hebrew University Business School and the Federmann Center for the Study of Rationality, The Hebrew University of Jerusalem, ORCID: 0000-0002-0978-322X; yochanan.bigman@mail.huji.ac.il

Corresponding author: Norbert Paulo (Norbert.paulo@gmail.com)

Abstract: The development of autonomous vehicles has stimulated research into preferences in sacrificial moral dilemmas. Studies such as the Moral Machine Experiment, which has attracted enormous attention in both philosophy and psychology, have been criticized for an important methodological flaw: forcing participants to choose between two options, like killing a man or a woman. It has been shown that many people actually prefer a third option, namely treating people "equally". While this is an important improvement, we argue that "equal" treatment can be understood in different ways. Instead of "equality", we propose to use the concept of impartiality and argue that impartiality needs to be complemented with action-guiding decision rules. To

support this conceptual point, we conducted a vignette study. Our results suggest that impartiality is indeed the main preference and that the most attractive decision rule to complement impartiality is not random choice, as many papers suggest, but inaction.

Competing interests statement: We have no conflicts of interest to disclose.

Acknowledgments: The authors thank Samuel Meyers for his assistance with statistical analysis.

YEB thanks the Israeli Science Foundation (Grant # 349/23) and the Asper Fund of the Hebrew University Business School for funding.

LK and NP gratefully acknowledge that this research is part of the project EMERGENCY-VRD and funded by dtec.bw – Digitalization and Technology Research Center of the Bundeswehr. dtec.bw is funded by the European Union – NextGenerationEU.

NP also gratefully acknowledges that this research is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 515122796; 532782236.

Keywords:

Autonomous vehicles, Trolley dilemma, equality, impartiality, inaction, randomness

1. Introduction: From forced preferences to "equality"

The emerging technology of autonomous vehicles (AVs) has stimulated research into people's preferences in sacrificial moral dilemmas where a choice needs to be made about who will live and who will die (Awad et al. 2018; Bonnefon, Shariff, and Rahwan 2016). While psychological research on moral dilemmas used to be considered overly hypothetical (e.g., Kahane 2015; Goodall 2016), the potential relevance of such dilemmas in cases of unavoidable accidents where AVs need to react increased the interest of the scientific community. But how should AVs be programmed for unavoidable accidents? Recently, several studies have investigated laypeople's moral preferences for such situations. Here, we focus on the consequences of methodological choices made in previous studies – specifically when addressing preferences for "equality" – and, in doing so, provide a more nuanced understanding of preferences in life-and-death decisions for AVs. We argue that the preference that has been discussed under the label "equality" is, at its core, a preference for impartiality (i.e., not considering personal characteristics) and that decision rules (such as random choice) are needed that are compatible with impartiality in order to reach a verdict about the programming of AVs. We test whether people are sensitive to different types of decision rules that are consistent with impartiality.

The largest, and most ambitious, study in this new line of research leveraging sacrificial moral dilemmas to the study of AV ethics is the Moral Machine Experiment (MME) (Awad et al. 2018) with its data set of 40 million decisions in ten languages from 2.3 million participants in 233 countries or regions. Awad et al. highlighted three dominant preferences for what they call a "universal machine ethic" (2018, 63): (1) sparing people before pets, (2) sparing as many lives as possible, and (3) sparing younger before older people. They also observed preferences for sparing the lives of the lawful as well as preferences concerning a variety of personal characteristics, such as sparing people with a higher social status, those who are fit (healthy), and women before men.

The MME and similar studies aim to inform the regulation of moral dilemmas with AVs although some question whether laypeople's preferences should inform policy on programming AVs to make life-and-death decisions (see, e.g., Nyholm 2018a; 2018b; Harris 2020; Savulescu, Gyngell, and Kahane 2021; Paulo 2023; Paulo, Möck, and Kirchmair 2023; Paulo and Kirchmair 2024). However, if laypeople's preferences are to be taken into account, the methodology for revealing such preferences should be as convincing as possible, which was not entirely the case with the MME.

In the MME, participants were presented with a static 2D rendering of a situation in which fatal accidents involving AVs are unavoidable. Visitors to the online platform were asked to decide what AVs should do in these situations by choosing between two accident scenarios in which, as a result of their choice, a certain person or group of people (or animals) is killed and another saved. Imagine the following situation: Two pedestrians – a woman and a child – are crossing the road. An AV is approaching fast and cannot stop in time. If it stays on course, it will kill the child. If it swerves, it will kill the woman. It cannot save both of them. In the MME, participants were asked to indicate whether the car should stay on course or swerve. It is crucial to note that only these two options are available.

Critics of the MME have pointed out that it really investigates forced preferences between either killing A or killing B, whereas alternative options might actually be preferred. This forced choice methodology lacks an option allowing participants to treat potential victims "equally", as Bigman and Gray (2020) called it. Bigman and Gray found that participants preferred AVs to be programmed to ignore personal characteristics like age and gender when they are given an explicit option of reporting such a preference. De Freitas and Cikara (2021) also criticize the forced choice design of the MME. They found that participants are morally outraged by programming which

leads to the deliberate killing of a specific social group (such as elderly or young people), even if this is not their favoured group. De Freitas and Cikara argue that people prefer randomness over the forced choice paradigm, suggesting that people are less morally outraged by AVs that kill certain groups by random choice than by ones that kill them deliberately. Bodenschatz, Uhl, and Walkowitz (2021) report similar findings, namely that participants prefer randomized outcomes in hypothetical and incentivized decision situations.

These preferences for not treating one group less favourably than another (Bigman and Gray 2020; De Freitas and Cikara 2021) are an important but limited first step in understanding people's preferences. We argue that there is a common core to all of these proposals – one that has not been recognized or made explicit yet. We argue that, in many cases, people prefer the potential victims of unavoidable accidents with AVs to be treated impartially, this being the common core that is important to participants. However, even these proposals fail to distinguish between a preference for impartiality and the decision rules that complement it, nor do they enable distinguishing between different decision rules.

2. Impartiality instead of "equality"

The argument suggested by critics of the MME that people have an "equality" preference sounds intuitively appealing. This idea, however, needs to be treated with care as it is not clear what "equality" means in the context of unavoidable accidents with AVs. As already mentioned, one possible interpretation involves the concept of randomness.

Consider the GENDER DILEMMA, a situation in which an AV is heading towards a woman. The only way to save her is to swerve so that a man is killed instead. However, such a dilemma could include a randomness option, where the AV decides randomly who to kill and who to

spare. Such a randomness option would be consistent with "equality" — no one is treated better or worse than the other.

In contrast, we suggest that "equality" is not the most helpful concept for understanding people's preferences in the context of sacrificial moral dilemmas, and suggest that impartiality is the more appropriate concept. According to the philosopher Bernard Gert, impartiality is a purely negative relation: an impartial choice is one in which a certain sort of consideration has no influence on the decision (Gert 1995; on the notion of impartiality more broadly, see Jollimore 2022). We think that Bigman and Gray 2020, De Freitas and Cikara 2021, and Bodenschatz, Uhl, and Walkowitz 2021 all have impartiality in Gert's sense at their core: the personal characteristics of potential victims must not influence the decision who should be killed and whose life is spared in AV dilemma situations.

Note that we are relying here on a notion of impartiality in which differences between individuals should not affect decisions. A less "pure" notion could also be considered in which only morally relevant differences should influence the decision. For example, a preference for doctors over criminals might be considered impartial because doctors are more beneficial to society. With such a broad notion of impartiality, the problem would be to decide which differences are morally relevant (for further discussion see Paulo and Kirchmair 2024).

De Freitas and Cikara argue that people prefer randomness over a forced choice paradigm because they are more morally outraged by AVs that kill discriminately than by those that kill indiscriminately (by random choice). Our suggestion is to reinterpret this finding such that people are less morally outraged by AVs killing randomly not because they find the random choice option attractive in itself. Rather, it is the only way to treat all potential victims impartially. People are less outraged by impartial killings than by killings in which certain considerations

(such as the personal characteristics of potential victims) influence the decision. What they find attractive about random choice is the impartiality inherent in it.

It is important to understand that while random choice and impartiality are related, they are different in relevant ways. In the GENDER DILEMMA, both random choice and impartiality would forbid deciding who to kill or save based on gender. However, they differ in what they suggest needs doing. Random choice would put the woman and man at equal risk of being killed. Impartiality, in contrast, does not suggest what needs doing; it does not determine who to kill in such situations. It merely defines how *not* to decide, namely by taking the personal characteristics of potential victims into account, leaving open the question as to what should be done instead.

To derive positive guidance on how to program AVs, we need to complement impartiality with action-guiding decision rules. Note that these decision rules must be consistent with impartiality. After all, there are a number of decision rules that are not impartial, such as the preferences (found in the MME) to favour young people at the expense of the elderly, or women at the expense of men. Although such decision rules provide clear guidance, they violate the basic idea of impartiality. As we show in the next section, random choice is not the only attractive candidate for a decision rule that is consistent with impartiality.

Summing up, impartiality does an important job. It tells you how *not* to decide, namely not to take the personal characteristics of potential victims into account. However, it needs to be complemented with action-guiding decision rules. This distinction between impartiality and decision rules clarifies the similarities and differences between the proposals by Bigman and Gray, De Freitas and Cikara, and Bodenschatz, Uhl, and Walkowitz and advances the debate.

3. Impartiality plus decision rules

If the preference that has been discussed under the label "equality" is better described as a preference for impartiality, decision rules need to be added to determine what AVs should be programmed to do. Without clearly distinguishing between impartiality and decision rules, Bigman and Gray (2020), for example, suggest preferences for saving more lives and a preference for inaction over action as part of the preference for "equality".

Consider the NUMBERS DILEMMA, a classical trolley scenario in which, for example, one course of action leads to the death of one person and another course of action leads to the death of three. Here, the decision rule to save as many lives as possible would lead to broadly utilitarian solutions. However, it would not determine any solution in cases in which both options lead to the same number of persons being killed like the GENDER DILEMMA.

For cases like the GENDER DILEMMA, another decision rule mentioned by Bigman and Gray might be considered, namely the rule for inaction over action (Ritov and Baron 1992; Spranca, Minsk, and Baron 1991; Yeung, Yay, and Feldman 2022). Combined with impartiality, in a situation where swerving would kill a man and staying on course a woman, this rule would lead to the woman being killed. This rule illustrates another difference between impartiality and "equality". While impartiality is a purely negative relation, the term "equality" might also have a normative-substantial element. For example, it can be understood as requiring some kind of equal distribution of outcomes, or equal chance at a certain outcome. In this normative-substantial understanding, it is not equal treatment to sacrifice the woman simply because she happens to be in the way of the AV. There is neither equal distribution of outcomes, nor equal chance at a certain outcome between the potential victims. Sacrificing the woman would, however, be impartial.

De Freitas and Cikara suggest using random choice as a decision rule (without distinguishing it from impartiality), as do Bodenschatz, Uhl, and Walkowitz. As already mentioned, in the GENDER DILEMMA, this would mean that the woman and the man have equal chances of being killed while in the NUMBERS DILEMMA, random choice as a decision rule would mean that the group of three has the same chance of being killed as the one individual.

Although random choice, saving as many lives as possible, and inaction over action are arguably the most obvious possibilities for decision rules that are compatible with impartiality, there are, of course, other possibilities, for example, alternating between staying on course and swerving, or always swerving right (or left). All of these decision rules can complement impartiality and provide direct guidance on how AVs should be programmed. However, adopting these different decision rules would lead to very different outcomes in the various dilemma scenarios with AVs, although they are all compatible with a preference for impartiality.

Note, further, that combinations of two or more decision rules are possible and might be warranted for certain situations. Suppose that an AV is about to crash into a child and its mother. There are only two options to prevent both dying. One is to swerve left, which would kill the child but save the mother. The other is to swerve right, which would kill the mother but save the child. In such a situation, the preference for impartiality tells us to disregard age as a factor when choosing between the two potential victims. But which decision rule should be adopted? The rule to save as many lives as possible would only tell the AV to swerve but not where to swerve; random choice would leave us with equal chances that the child dies, the mother dies, or both die; the rule for inaction over action would lead to both dying. Here, combinations of decision rules might be warranted, such as one that prevents the unnecessary death of both potential victims. For example, this could be achieved by combining random choice with the rule to save as many

lives as possible such that the set of options for random choice is limited to killing the child or killing the mother.

As we have shown, "equality" is not the most appropriate concept for understanding people's preferences in the context of unavoidable accidents with AVs. Instead of "equality", we propose to use the concept of impartiality (as a negative relation in Gert's sense). Impartiality then needs to be complemented with an action-guiding decision rule that is consistent with impartiality.

It is important to incorporate this conceptual clarification into an empirical investigation of preferences for how AVs should be programmed for sacrificial dilemma situations. We need to test whether laypeople really prefer impartiality and, if so, which decision rule they prefer to complement impartiality in which situation. Our study is designed to test whether prior results are best explained by a specific preference for randomness, inaction or alternation, or a more general preference for impartiality.

In addition, we investigate how different direct decision rules affect such a preference and whether the type of dilemma (e.g., the AGE vs. GENDER DILEMMA) makes a difference.

This paper makes two contributions to the literature. First, by drawing on work in philosophy (Gert 1995), we provide a more precise concept for a third option in sacrificial moral dilemmas. Instead of "equality", we suggest that impartiality is more appropriate. Second, because impartiality cannot be tested directly but only in combination with decision rules, we tested whether participants show similar preferences for decision rules that satisfy the condition of impartiality, see Figure 1.

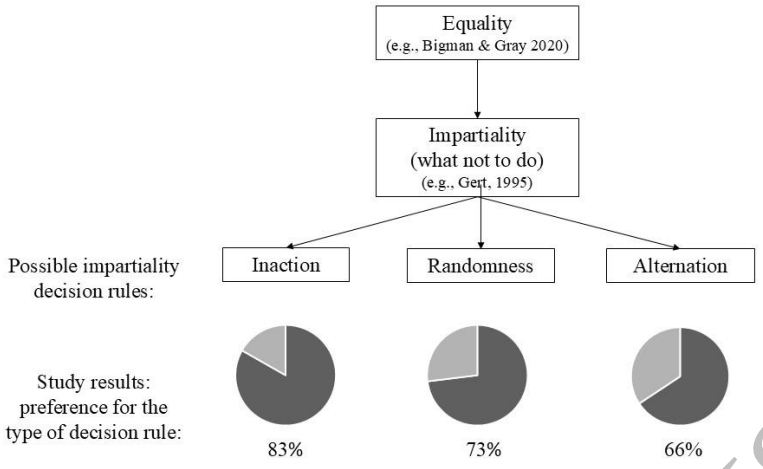

Figure 1. The two main contributions of the paper. First, replacing the ill-defined concept of "equality" used in previous research and use the concept of impartiality instead. Second, testing whether people are sensitive to types of decision rules that are consistent with impartiality.

4. Study

We conducted a vignette study to examine people's preferences for different types of decision rules that all complement impartiality understood as a negative relation. In this study, we focused on three dilemmas where, due to faulty brakes, an AV needs to decide who to kill and who to save: the AGE DILEMMA (whether they prefer to spare the life or a child or an elderly person), the STATUS DILEMMA (whether to spare the life of a doctor or a homeless person), and the GENDER DILEMMA (whether to spare the life of a man or a woman). We chose these dilemmas because previous research reports that the strength of people's preferences for impartiality varies between these dilemmas (e.g., Bigman and Gray 2020). To obtain comparable

results with previous work, we closely followed the design of the MME and the Bigman and Gray study.

We created three versions for each type of dilemma. In each, the first two options were the same (e.g., "save the child", "save the elderly person"). The third option (the impartiality option) was framed as one of three decision rules that are consistent with impartiality, namely randomness, alternation, or inaction. All participants rated their preferences for all nine scenarios.

4.1. Methods

4.1.1. Pre-registration and data availability. We registered the study at https://aspredicted.org/J79_689. Data and code are available at: https://osf.io/u5ynw/?view_only=0deff5a7248c4fd683c45d41229a8d06.

4.1.2. Participants. We sampled 102 participants (Age: $M = 37.11$, $SD = 10.98$; Gender: 48% male, 52% female, 0% other/prefer not to answer) on 27 November 2023 using SoSci and prolific.

4.1.3. Procedure. We used a 3 (dilemma type: age, gender, and social status) x 3 (type of decision rule: inaction, random choice, and alternation) within-subject design.

Participants first completed an attention check where they saw a simple maths question but were asked to select the third option, although it was false. After the attention check, participants read a short introduction to the moral dilemmas that self-driving cars might face (see appendix for full text).

The dilemmas appeared in the following format, this being an example of the SOCIAL STATUS dilemma, in the decision rule=alternation version (see appendix for all dilemmas):

"Imagine an accident situation in which a self-driving car has only two options. It can either run over a doctor or a homeless person.

Your task is to decide how to program self-driving cars to act in such situations.

What should self-driving cars be programmed to do?

1. To run over the doctor (killing the doctor).
2. To run over the homeless person (killing the homeless person).
3. To alternate between running over the doctor (killing the doctor) and running over the homeless person (killing the homeless person).

Although the answer options may look similar, they are not identical. Please pay careful attention to the options available in each condition. Thank you very much!"

The dilemmas, and the options, were presented in random order. Participants finally reported their age and gender.

4.2. Results

4.2.1. Analytical approach

To test whether differences in percentages of participants who chose the impartiality option are significant, we conducted a statistical test of significance rather than just examining means. Because our dependent variable is binary (impartiality or not), we used logistic regression. To account for the nested nature of our data, as each participant provided preferences for nine scenarios (three versions for three dilemmas), we used mixed effect regression, with random intercepts for each participant (the model with random slopes per participant did not converge). We used dummy coding for the decision rule and dilemma type with inaction and gender as the reference conditions respectively. Our predictors were decision rule and dilemma type (as we had no theory about the interaction). The percentages of preference for impartiality in each category

are presented in Table 1. Full data about the percentage of participants who preferred the impartiality option in each of the nine versions are presented in Figure 2.

4.2.2. Preference for impartiality

We found a significant preference for impartiality across all decision rules and dilemma types such that participants preferred the impartiality option 68.08% of the time, *Intercept*=4.83, $p < .05$, OR = 125.39, $CI_{.95}$ [48.17, 326.36]. (Age might be considered an exception; see the discussion below.)

4.2.3. Preference for type of decision rule

We also found a significant preference for type of decision rule, such that participants preferred inaction (76.14%) more than both alternation (61.11%; B = -1.59, $p < .05$, OR = 0.25, $CI_{.95}$ [0.12,0.36]) and random choice (66.99%; B = -1.00, $p < .05$, OR = 0.37, $CI_{.95}$ [0.21, 0.63]).

4.2.4. Preference for dilemma type

We found a significant effect for dilemma type, such that the preference for impartiality was higher for gender (89.54%) than for social status (73.53%; B = -1.93, $p < .05$, OR = 0.15, $CI_{.95}$ [0.08, 0.27]) but much lower for age (41.17%; B = -4.60, $p < .05$, OR = 0.01, $CI_{.95}$ [0.00, 0.02]).

Category	% picking impartiality
Overall	68.08%
Rule [inaction]	76.14%
Rule [alternation]	61.11%

Rule [random]	66.99%
Dilemma type [gender]	89.54%
Dilemma type [status]	73.53%
Dilemma type [age]	41.17%

Table 1. Preferences for impartiality for each condition.

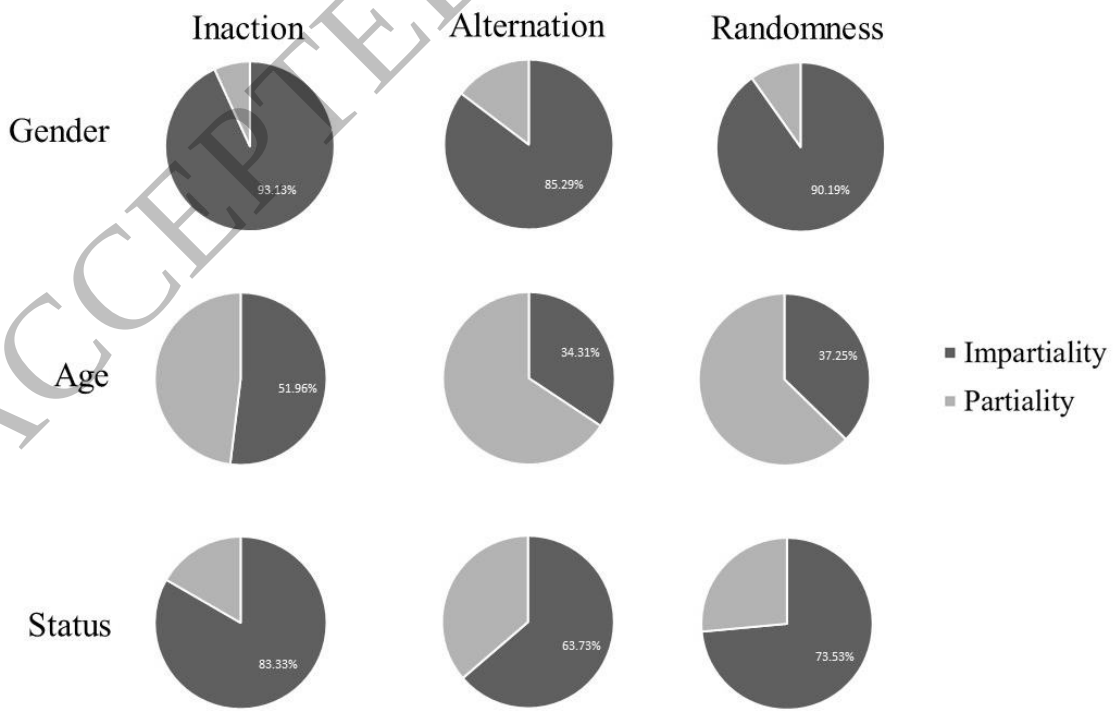

Figure 2. Percentages of participants who chose the impartiality option in each of the nine scenarios

5. Discussion

The criticism of the study design in the MME (Awad et al. 2018) and similar studies regarding the forced choice methodology (between two different individuals with different personal characteristics) is justified. Most people do not actually favour prioritizing individuals with certain personal characteristics but want to treat them "equally" (Bigman and Gray 2020; De Freitas and Cikara 2021; Bodenschatz, Uhl, and Walkowitz 2021).

As intuitively compelling as this may be, "equality" can be understood in very different ways. First, we suggest that this third option should not be conceptualized as "equality" but is best understood as impartiality. The main reason is that impartiality is a purely negative relation leading to certain considerations simply being disregarded when making a decision (Gert 1995). Therefore, when faced with a moral dilemma, we also need decision rules that tell us what to do. Since there are many different options for such rules, it is important to explore these different options.

Our results show that not all decision rules are equally attractive. It is, therefore, important to know which ones are consistent with impartiality and to use them accurately in empirical studies. It makes a difference whether respondents have an impartiality option in the form of a decision rule on inaction, random choice, or alternation. In terms of how to inform the debate on regulating moral dilemmas involving AVs, it is important to consider that, despite such

impartiality options, many individuals will still prefer one type of personal characteristic over another.

The most attractive decision rule seems to be inaction. This is understandable as it does not involve a deliberate decision to kill someone. Moreover, different personal characteristics are treated differently. For example, many more participants are willing to protect young people at the expense of the elderly than those willing to protect people with a presumably higher social status, like doctors, at the expense of those with a lower social status, like the homeless. We also note that in some cases people might reject impartiality, such as in cases of jaywalking where some of the people might be personally responsible for causing the dilemma. Future research should systematically test these cases.

We note that our results, specifically with preferences towards age, differ considerably from previous findings. While in our sample only 41.17% of participants preferred age impartiality (and 58.83% preferred to save the young), in Bigman and Gray (2020), 77.3% preferred "equality" between young and old in Study 1 and 61.1% in Study 2. The different framing of the "impartiality" option might explain this: in our study, participants were always asked about impartiality indirectly through answer options that featured one of the decision rules that are consistent with impartiality while in Bigman and Gray (2020) participants were directly asked if they preferred to treat the potential victims "equally" (leaving open what "equally" means in the scenario concerned). In addition, our own research suggests that the different decision rules matter. For these reasons, it is difficult to compare our results on age with those in previous studies (for further discussion see Paulo and Kirchmair 2024).

Further research is needed to explore in more detail which decision rules laypeople prefer for which dilemmas with AVs. We have shown that the criticism of the forced choice design of the

MME was correct but that it is not enough to simply offer a vague third option. It is important to provide test participants with clearly formulated decision rules.

6. Bibliography

Awad, Edmond, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon, and Iyad Rahwan. 2018. "The Moral Machine Experiment." *Nature* 563 (7729): 59–64. https://doi.org/10.1038/s41586-018-0637-6.

Bigman, Yochanan E., and Kurt Gray. 2020. "Life and Death Decisions of Autonomous Vehicles." *Nature* 579 (7797): E1–2. https://doi.org/10.1038/s41586-020-1987-4.

Bodenschatz, Anja, Matthias Uhl, and Gari Walkowitz. 2021. "Autonomous Systems in Ethical Dilemmas: Attitudes toward Randomization." *Computers in Human Behavior Reports* 4 (August):100145. https://doi.org/10.1016/j.chbr.2021.100145.

Bonnefon, Jean-François, Azim Shariff, and Iyad Rahwan. 2016. "The Social Dilemma of Autonomous Vehicles." *Science* 352 (6293): 1573–76. https://doi.org/10.1126/science.aaf2654.

De Freitas, Julian, and Mina Cikara. 2021. "Deliberately Prejudiced Self-Driving Vehicles Elicit the Most Outrage." *Cognition* 208 (March):104555. https://doi.org/10.1016/j.cognition.2020.104555.

Gert, Bernard. 1995. "Moral Impartiality." *Midwest Studies In Philosophy* 20 (1): 102–28. https://doi.org/10.1111/j.1475-4975.1995.tb00307.x.

Goodall, Noah J. 2016. "Away from Trolley Problems and Toward Risk Management." *Applied Artificial Intelligence* 30 (8): 810–21. https://doi.org/10.1080/08839514.2016.1229922.

Harris, John. 2020. "The Immoral Machine." *Cambridge Quarterly of Healthcare Ethics* 29 (1): 71–79. https://doi.org/10.1017/S096318011900080X.

Jollimore, Troy. 2022. "Impartiality." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2022. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2022/entries/impartiality/.

Kahane, Guy. 2015. "Sidetracked by Trolleys: Why Sacrificial Moral Dilemmas Tell Us Little (or Nothing) about Utilitarian Judgment." *Social Neuroscience* 10 (5): 551–60. https://doi.org/10.1080/17470919.2015.1023400.

Nyholm, Sven. 2018a. "The Ethics of Crashes with Self-Driving Cars: A Roadmap, I." *Philosophy Compass* 13 (7): e12507. https://doi.org/10.1111/phc3.12507.

———. 2018b. "The Ethics of Crashes with Self-Driving Cars: A Roadmap, II." *Philosophy Compass* 13 (7): e12506. https://doi.org/10.1111/phc3.12506.

Paulo, Norbert. 2023. "The Trolley Problem in the Ethics of Autonomous Vehicles." *The Philosophical Quarterly* 73 (4): 1046–66. https://doi.org/10.1093/pq/pqad051.

Paulo, Norbert, and Lando Kirchmair. 2024. *Moral Dilemmas Involving Self-Driving Cars: How to Regulate Them and Why Your Opinion Matters*. New York, NY: Routledge.

Paulo, Norbert, Leonie Möck, and Lando Kirchmair. 2023. "The Use and Abuse of Moral Preferences in the Ethics of Self-Driving Cars." In *Experiments in Moral and Political Philosophy*, edited by Hugo Viciana, Antonio Gaitán, and Fernando Aguiar, 290–309. London: Routledge.

Ritov, Ilana, and Jonathan Baron. 1992. “Status-Quo and Omission Biases.” *Journal of Risk and Uncertainty* 5 (1): 49–61. https://doi.org/10.1007/BF00208786.

Savulescu, Julian, Christopher Gyngell, and Guy Kahane. 2021. “Collective Reflective Equilibrium in Practice (CREP) and Controversial Novel Technologies.” *Bioethics* 35 (7): 652–63. https://doi.org/10.1111/bioe.12869.

Spranca, Mark, Elisa Minsk, and Jonathan Baron. 1991. “Omission and Commission in Judgment and Choice.” *Journal of Experimental Social Psychology* 27 (1): 76–105. https://doi.org/10.1016/0022-1031(91)90011-T.

Yeung, Siu Kit, Tijen Yay, and Gilad Feldman. 2022. “Action and Inaction in Moral Judgments and Decisions: Meta-Analysis of Omission Bias Omission-Commission Asymmetries.” *Personality and Social Psychology Bulletin* 48 (10): 1499–1515. https://doi.org/10.1177/01461672211042315.