Moving away from The Trolley Problem - Methodological improvements due to ADC Model operationalized virtual traffic vignettes

Low-level Autonomous vehicles (AVs), merely assisting human drivers, are already being sold in the U.S., and the deployment of higher-level AVs in the future is likely [1]. While behavior guidance systems (the hardware and software architectures that enable intelligent motion planning) are rapidly advancing, the rules that underpin these systems are technically and ethically insufficient for the feasible incorporation of higher-level AVs, which are entirely autonomous [2]. Since these rule sets underscore many ethical issues, including safety and accountability, the industrial sector is interested in addressing this limitation [3]. Intending to minimize harm and safeguard companies, some moral philosophers are contributing to solving this rule set issue by utilizing the intuitions guiding human decision-making during traffic dilemma experiments. The hope is to inform these systems of how human drivers, a proven system, navigate the abstract moral layer of these complex circumstances.

Numerous studies have addressed how human decision-making can inform the moral actions of AVs using various methodologies. The most prominent experimental approach is some variation of the trolley problem thought experiment, a high-stakes sacrificial dilemma where both options typically involve the hypothetical death of pedestrians or passengers [3]. This setup offers many experimental pros for assessing how people respond to morally taxing traffic dilemmas. Take, for instance, the highly publicized Moral Machine experiment, a large-scale online survey that evaluates global preferences (the salient characteristics of the scenario that influence the participants' decisions) [4]. This design is simple, repeatable, offers explanatory power, and evaluates these salient traffic dilemmas, which at face value provides everything necessary to address this rule set issue.

While Moral Machine and similar trolley problem-based experiments are a useful starting point, this group has identified three aspects of the typical paradigm worth altering to better inform the rule sets necessary for deploying higher-level AVs. (1) The trolley-problem paradigm is appealing because it has provided a modern battleground for the consequentialists, pushing utilitarian greater good ideals, and the deontologists, basing their moral stance on rule-based rationales, but this ignores important morally salient information captured by virtue ethics. (2) Actual traffic dilemmas also involve low-stakes moral salient events, which must be considered alongside the high-stakes counterparts for a robust rule set. (3) Any uni-cultural rule set will have inherent normative biases, opening the need for language-independent stimuli.

Our group's original solution to these three deficiencies is to create virtual traffic dilemmas operationalized based on the Agent-Deed-Consequence model of moral judgments [5]. This ethical model incorporates the three main ethical pillars of moral philosophy: virtue ethics, deontology, and utilitarianism, operationalized into relevant moral stimuli [6]. Secondly, while higher-stakes versions of scenarios are possible, this model focuses primarily on non-fatal accidents and traffic infractions [7]. Finally, the multimodal and dynamic stimuli available in virtual development have the potential to be completely language-independent [8], meaning there will be no loss of context during survey transitions. In sum, creating an experimental design that maintains the trolley problem core while eliminating a glaring issue, adding nuance to the results, and broadening the participant pool.

References:

1. Canis, B. (2019). Issues in autonomous vehicle testing and deployment (No. R45985). Congressional Research Service.

2. Sharma, O., Sahoo, N. C., & Puhan, N. B. (2021). Recent advances in motion and behavior planning techniques for software architecture of autonomous vehicles: A state-of-the-art survey. Engineering applications of artificial intelligence, 101, 104211.

3. Martinho, A., Herber, N., Kroesen, M., & Chorus, C. (2021). Ethical issues in focus by the autonomous vehicles industry. Transport reviews, 41(5), 556-577.

4. Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.F., & Rahwan, I. (2018). The moral machine experiment. Nature, 563(7729), 59-64.

5. Dubljević, V. & Racine E. (2014): The ADC of Moral Judgment: Opening the Black Box of Moral Intuitions with Heuristics about Agents, Deeds and Consequences. Target Article: AJOB – Neuroscience, 5 (4): 3-20, (2014).

6. Dubljević, V., Sattler, S., & Racine E. (2018): Deciphering moral intuition: How agents, deeds and consequences influence moral judgment, PLOS One, doi: 10.1371/journal.pone.0204631.

7. Dubljević, V. (2021): Toward Implementing the ADC Model of Moral Judgment in Autonomous Vehicles. *Sci Eng Ethics* (2020). https://doi.org/10.1007/s11948-020-00242-0

8. Sütfeld, L. R., Ehinger, B. V., König, P., & Pipa, G. (2019). How does the method change what we measure? Comparing virtual reality and text-based surveys for the assessment of moral decisions in traffic dilemmas. PloS one, 14(10), e0223108.