

IS DRIVING SIMULATION A VIABLE METHOD FOR EXAMINING DRIVERS' ETHICAL CHOICES? AN EXPLORATORY STUDY

Anuj K. Pradhan¹, Heejin Jeong², Brittany Ross²

¹University of Massachusetts Amherst, Amherst, MA, USA

²University of Michigan, Ann Arbor, MI, USA

Email: anujkpradhan@umass.edu

Summary: Advanced vehicle technologies promise improved road safety but may still be subjected to situations where choices have to be made regarding safety impact to other road users. There is debate about the principles that should guide the programming of choices into automation algorithms, and an acknowledgment that choices made by automation may be subject to more scrutiny than those by humans. To better understand the landscape of decisions that human drivers encounter, it is important to examine the rationale, calculus, and motivations behind such choices. While there are various methods to examine human decision making, doing so in an ecologically valid manner is challenging, especially in this context of driving. To that end, this study was conducted to examine if driving simulation could help understand drivers' ethical choices. Participants drove a route in a driving simulator that was programmed to end in a crash situation, one that placed the driver in a position of choosing between two crash outcomes. Participants were asked, after the fact, about their perceptions of the simulation and their decisions. Results indicate that drivers generally accepted simulation as realistic, but their post-experiment choices did not align with their actual decisions during the drive. Findings may have implications for the experimental study of ethical behaviors.

BACKGROUND

Motor vehicle crashes are a major source of injuries and fatalities in the United States and a recognized public health problem. While automated vehicles have the potential to improve safety, it remains the case that automated vehicles may still face situations where decisions have to be made to prevent a crash, or if that fails, to undertake actions to minimize consequences. In situations where a crash is unavoidable, algorithmic choices will ultimately result in a particular outcome, based on pre-programmed rules or heuristics (Gerdes & Thornton, 2015). However, opinions regarding the 'best course of action' may differ considerably based on one's motivation, cost calculation, or moral or ethical stance. Indeed, situations could arise that result in choices between an individual's interest (safety, comfort, convenience), other road users' interests, or societal/community interests (Fleetwood, 2017). Often, these situations are not explicitly anticipated and programmed for, but rather algorithms are informed by rule-based or outcome-based programming. Thus, algorithms that ultimately make decisions could be programmed differently depending on principles regarding 'minimum consequence', principles that may not necessarily have broad consensus regarding ethics or morality. Previous simulator studies found that ethical decisions were mostly made based on a utilitarian manner, or a desire for a utilitarian result, with less concern about other details (Faulhaber et al., 2017; Frison, Winterberger, & Riener, 2016; Skulmowski, Bunge, Kaspar, & Pipa, 2014)

Ethical issues of automated vehicles may have a large impact on different groups (Lin, 2015), including manufacturers, the public, insurance companies, and so on. Automated vehicles must continually manage risks without human oversight, and when crashes are unavoidable automated vehicles have the potential to calculate the best way to crash using a combination of safety, the likelihood of the outcome, and certainty in measurements much faster than humans can (Goodall, 2014). These automated vehicles are thus tasked with balancing safety, mobility, convenience, ethics, as well as legalities. However, there are currently no standards and models to guide the programming of normative human ethics into the decision systems of automated vehicles. Neither is it well established what the outcomes are of the diverse ethical principles relating to situations involving automated vehicles. Current research seems to agree that there is no obvious way to code human morals into software that controls automation (Goodall, 2014), especially given that understanding or predicting humans' ethical choices itself is challenging.

OBJECTIVE

One approach to studying ethical decisions is by surveying people about their responses to a hypothetical situation. Such surveys have indeed been conducted and do yield rich information about humanity's ethical decisions in cases of forced choice situations (Awad et al., 2018). However, these surveys have limited external validity: a survey about the crash imminent situation when a respondent has the luxury to provide a considered response may be different from having to respond immediately when actually encountering such a situation. An alternative is to empirically examine such human reasoning. But there exist safety and ethical challenges in doing so. However, experimental methods to empirically study moral reasoning may be viable, if the approaches satisfy the dual conditions of ecological validity and safety. Empirical information from drivers has long been used to study their behaviors. One approach towards exploring ethical choices may thus be to study human responses when actually in such situations, especially where any choice made by the driver could result in a range of crash and injury severity. Such data cannot be obtained safely from real-world driving. However, it may be possible in a simulated environment. The objective of this study was therefore to explore this issue of examining ethical reasoning experimentally. More specifically, this study examined if realistic driver responses to crash-imminent scenarios could be obtained using simulation, and whether the actual driver decisions were aligned with drivers' responses to surveys.

METHODS

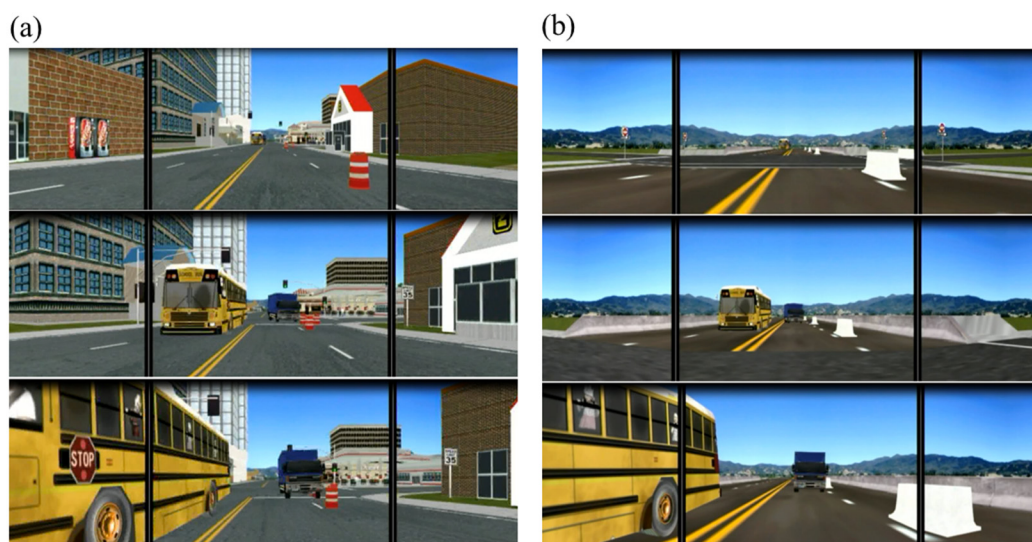
Overview

Participants manually drove a simulated route that ended in a crash. This crash was programmed so that the driver could not maneuver out of it and had to decide how to crash, with implications for harm to other road users. (The vehicle was not automated in this study, despite implications for vehicle automation.) The study was approved by the university's institutional review board.

Driving Simulation, Crash Scenario, & Simulated Drive

A fixed-base simulator (Realtime Technologies Inc), comprising a full vehicle in a dedicated lab space was used for the study. The virtual world was projected on three screens 16 feet in front of the driver (120° field of view) and a rear screen 12 feet away (40° field of view). Candidate

scenarios representing possible realistic crash-inevitable situations were identified and discussed with industry partners. The intent was to select a scenario where the driver had to make a quick decision on whether to crash in a manner that could result in minimal harm to all other road users involved. However, any choice would result in certain virtual ‘harm’ to the driver. Several guiding principles were used to identify scenarios: 1) realism; 2) relevance to automated vehicles; and, 3) limited response choices for the driver, each of which resulted in crashes that differed in critical ways in terms of likely severity of injuries and who was placed at greatest risk. Of multiple candidate scenarios, one was selected and programmed into two drives, Drive A and Drive B. Each drive ended in a similar crash scenario, but in different roadway types. Specifically, Drive A was in a city intersection and Drive B was on an overpass. The scenario was a ‘sudden-reveal’ involving the participant, a truck, and a school bus. Panels a & b in Figure 1 illustrate the scenarios in Drives A and B (respectively) via three images of the driver’s view showing the evolving crash in the forward direction as he/she approached the scenario.



**Figure 1. Driving Scenarios: (a) Drive A - city intersection, (b) Drive B- overpass
(from driver’s point-of-view)**

The scenario played out as follows: drivers saw a school bus in the opposite lane, heading in the opposite direction. A semi-truck that was closely following the bus then emerged from behind the bus (to the right from the bus from the participant’s point of view) ostensibly to pass the bus. This would occur just as the bus neared the participant, such that the oncoming semi would veer into the participant’s lane resulting in an imminent head-on collision between the driver and the semi. Barrels, or jersey barriers, were located to the right of the driver’s path. Given the speeds involved (~30 mph) and the last-minute nature of the semi’s intrusion, the situation evolved very quickly, thereby limiting the choices available to the driver. The scenario was also designed so that at the critical moment when the semi loomed towards the participant, the participant’s vehicle brake stopped working, so the participant had to choose to swerve. This left the participant with the options of not steering and being hit head-on by the truck, swerving to the right into the barriers/barrels to try to avoid all vehicles but likely injuring themselves, or swerving to the left and likely hitting the school bus and injuring themselves.

Participants and Procedure

Sixteen licensed drivers between 30-55 years (9 males, 7 females) were recruited. They were not sensitive to motion sickness, not professional drivers, not pregnant, did not have a history of depression or Post Traumatic Stress Disorder, and did not have a history of a major car crash. After signing an informed consent form, participants were given a 5-minute practice drive in the simulator. Since participants' expectations could change after exposure to one crash, each participant was only exposed to one scenario. Also, because the scenarios were crashes, each drive only included a single scenario. Therefore, each participant drove one experimental drive, 10-15 minutes long, containing one scenario. The first 10-12 minutes of the drive comprised an uneventful route to allow the participant to become comfortable driving the simulated vehicle in the virtual environment, reduce over-vigilance, and to lull them into a sense of regular, unremarkable, everyday driving. The drive included typical roadway features such as straight and curved, expressway, rural, and urban roads. The driver then encountered one crash scenario, immediately after which the experimental driving session ended. After the drive, an in-person structured interview was conducted using a questionnaire designed to probe participants about their experience. The last two questions in this session were preceded by the participants viewing a pre-recorded example video of the scenario (minus crash response) to illustrate the crash they had just experienced. Participants were not informed that the purpose of the study was to explore ethical decisions until after the study visit ended. Participants were paid a \$20 incentive. The protocol took about one hour.

Outcome Measures and Analyses

Outcome measures included interview responses and driver responses in the simulator. Responses to interviews were transcribed and studied to identify key themes and responses for each participant. Drivers' responses to the crash were coded from drive videos and simulator data. Data analyses were conducted to characterize responses to interview questions, and to examine whether drivers' responses to the scenario were aligned with the surveyed responses.

RESULTS

The outcomes can be broadly categorized into: those regarding participants' perception of the simulation, including realism; and participants' attitudes towards the crash scenario and calibration of their actions.

For the first, participants generally reported experiencing and interpreting the simulation as per the experimental intent and design. While describing the crash, all participants described it as sudden, quick, or unexpected, and that the "truck swerved towards" them. However, none characterized the situation explicitly as one that forced an ethical choice. In terms of the realism of the experience, based on three questionnaire items (i.e., "*Did the event at the end of the drive give you an adrenalin rush?*"; "*How realistic did it [the crash] feel?*"; and "*Do you feel your response was similar to what you would have done in the real world in the same situation?*") that were combined to form a binary determination of the participants' experience, 87.5% (14 participants) reported the crash experience as feeling realistic and 12.5% (2 participants) reported it as unrealistic.

For the second category, a number of findings can be reported.

Attitude and feeling after crash. Three items from the post-drive questionnaire were coded for three levels of perception/affect: positive, negative, or neutral. These were then combined to form a composite response. 68.8% (11 participants) had responses that were negative, and 31.3% (5 participants) had responses that seemed to indicate a neutral attitude/affect towards the crash event. No participants reported positive (or related) feelings or attitudes after the crash scenario.

Would participants have responded differently. To examine self-perceptions of their actions, participants were specifically asked if they wished they had responded differently while driving. 37.5% (6 participants) said they did wish they had responded differently, with most reporting wishing they had driven slower or paid more attention. 62.5% (10 participants) did not wish they had done anything different. Most of these participants seemed to have felt there was no other option they could have chosen, and that the crash was inevitable.

Response options available & ethical selection. Prior to the final questions in the post-interview, participants viewed a pre-recorded video of the crash. They were then asked to (1) list all possible response options available to the driver; and, (2) pick one response that they personally considered the most ethical option. The participants collectively listed eight actions/combinations that were available to the driver: *brake*; *swerve right*; *swerve left*; *brake & swerve right*; *brake & swerve left*; *no-response*; *accelerate*; and, *try to attract truck's attention (horn, etc.)*. Of these, in response to the second question, 37.5% chose '*swerve right*', 25% chose '*brake*', 18.8% chose '*swerve left*', 12.5% chose '*brake and swerve left*', and 6.3% chose '*brake and swerve right*' as the most ethical option. No participants selected the other actions (accelerating, hitting truck head on, and try to get truck's attention) as ethical choices.

Actual driving responses. Actual driver responses obtained from video and simulator data show that 56.3% (9 participants) steered left, 18.8% (3 participants) steered right, and 25% (4 participants) did not swerve at all. In terms of braking, 87.5% (14 participants) depressed their brakes, and 12.5% (2 participants) did not. Since the experiment was designed such that the brakes 'failed', even if a participant stepped hard on the brakes, the vehicle did not slow. Given this fact, and the fact that the braking attempt actions were close to universal, subsequent coding of driver behaviors did not take into consideration the braking behaviors.

Action vs choices. Using the above data, a final analysis was carried out to study the calibration of participants' choices between: ethical response when surveyed, i.e., when they could make the choice without a time pressure after viewing a video; and, the actual driving choices that were made when participants were immersed in the situation with time and other pressures. First, participants' responses were coded according to their chosen 'most ethical' action in the interview question, viz., (a) whether the action would result in harm to the participant himself/herself, or (b) result in harm to the school-bus (and to the participant). Second, participants' actual actions in the simulator were coded according to the final outcome, i.e., (a) if the action resulted in a crash that led to harm to the participant alone, or (b) if the action resulted in a crash that harmed the school bus (and to the participant). The following coding guidelines were used: If the participant turned (or chose as the most ethical option to turn) left into the school bus, the action was coded as harmful to the school bus. If the participant turned (or chose

as the most ethical option to turn) right, or to not turn, or to only the brake, then the action was coded as harmful to self only.

The result of this coding showed that, when surveyed, 68.75% of the participants chose options that would avoid harm to the school bus. However, when actually behind the wheel, only 43.75% of the drivers actually undertook actions that avoided harm to the school bus. Further analyses for individual participants show that only half of the participants had the same actual behavior that they chose as the most ethical during the interview. This data suggests a mismatch in participants' intent when surveyed, as opposed to their actions when actually in a situation. A McNemar's Chi-squared test with continuity corrections comparing 'choice' versus 'actual behaviors' for 'harm to the bus' and 'harm to self' in a 2×2 contingency table, however, showed no significance in the change of actions from 'surveyed' to 'actual' (McNemar's Chi-squared = 1.125, $df = 1$, $p = .2888$). This lack of significance may be explained by the low sample size for the study with only 16 matched-pairs for the contingency table. Nonetheless, it is still interesting to see the difference in magnitude of intended and actual behaviors, suggesting that responses made when not actively involved in a situation or choices made with the luxury of time to reflect on consequences, may be biased in terms of probing for moral reasoning.

CONCLUSION

The results of this study show that drivers generally understood the simulation and accepted the simulation and the crash as realistic, and as per design, with accurately corresponding feelings/perceptions generated post-crash. Despite this, there seems to be a mismatch between driver action when in the moment and driver intent when not actively driving. Although not a statistically significant result, it is interesting to see that half of the drivers reported, when surveyed, that they would undertake a reaction opposite of what they actually did. One reason for this disconnect could be because drivers in the simulator drove differently than they would have in the real world due to differences between simulated and real driving. However, the fact that a majority of participants attested to the realism of the simulation indicates that this may not be the case. Another reason could be that actions taken when immersed in a situation are different from behaviors one thinks one would have when responding to a hypothetical survey question.

When asked about the most ethical choice, drivers generally chose one that did not result in harm to the bus. It is uncertain if the safety of the bus was indeed taken into consideration when making these choices. In fact, it is uncertain if the participants did notice the school bus for what it was as the scenario materialized. One could assume that given the rapidly emerging crash scenario, all of the driver's attention could potentially have been focused on the oncoming semi and they had little awareness of the surrounding traffic elements. Nonetheless, when comparing people's actions with their choices, behaviors were not aligned, meaning, they hit the bus. It is not possible to state with certainty if this mismatch manifested because people thought it would be safer to hit the bus, or if it was a reflexive and unplanned reaction. The evidence, therefore, suggests that drivers may make ideal choices in a hypothetical situation when not faced with the actual task of maneuvering a vehicle under time and outcome pressure. When actually in such a situation, one's actual behaviors may be different from their hypothetical choices. This particular result should be examined further, especially given that research that relies on surveys of drivers have contributed some evidence to the field about driver ethical choices.

There are a number of limitations in this study. An important one is the modest sample size. Given the exploratory nature of the study, data were collected from just sixteen participants. It is likely that the study was underpowered for a significant result regarding the mismatch between choice and action. Another is the use of a fixed-base simulator. A motion base may have increased realism enough to differently influence driver reactions. In fact, during debrief, most participants specifically identified the lack of motion as being an element that hindered realism.

Despite limitations, this study provides important findings regarding the use of driving simulation for studying drivers' rationale for moral choices. Prior research has not attempted to use driving simulation to study driver reactions to situations in which a crash is imminent and unavoidable in order to determine what natural responses would be on the part of experienced drivers; thus, evaluating drivers' ethical choices in vehicle automation is timely and necessary. This study could be a first step in a process to understand the choices people make and their reasons for those choices, and an initial step in investigating the ethical issues associated with vehicle automation and their involvement in imminent crashes. Progress in this area will require manufacturers, designers, and policymakers to understand the social and behavioral implications of vehicle automation, including ethical dilemmas. Simulation could provide the ability to assess a wide range of safety-critical situations, such as pedestrian and bicycle interactions, before putting anyone at risk, and to understand what normative behaviors are in such situations.

ACKNOWLEDGEMENTS

This study was funded by the University of Michigan Mcity.

REFERENCES

- Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., ..., & Rahwan, I. (2018). The Moral Machine Experiment. *Nature*, 563 (7729), 59.
- Faulhaber, A. K., Dittmer, A., Blind, F., Wächter, M. A., Timm, S., Sütfeld, L. R., ... & König, P. (2018). Human Decisions in Moral Dilemmas are Largely Described by Utilitarianism: Virtual Car Driving Study Provides Guidelines for Autonomous Driving Vehicles. *Science and Engineering Ethics*, 1-20.
- Fleetwood, J. (2017). Public Health, Ethics, and Autonomous Vehicles. *American Journal of Public Health*, 107(4), 532-537.
- Frison, A.K., Wintersberger, P., & Riener, A. (2016). First Person Trolley Problem: Evaluation of Drivers' Ethical Decisions in a Driving Simulator. In *Adjunct Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (AutomotiveUI '16 Adjunct). ACM, New York, NY, USA, 117-122.
- Gerdes, J. C., & Thornton, S. M. (2015). Implementable Ethics for Autonomous Vehicles. In *Autonomes Fahren* (pp. 87-102). Springer Vieweg, Berlin, Heidelberg.
- Goodall, N. J. (2014). Ethical Decision Making During Automated Vehicle Crashes. *Transportation Research Record*, 2424(1), 58-65.
- Lin, P. (2015). Why Ethics Matters for Autonomous Cars. In *Autonomes Fahren* (pp. 69-85). Springer Vieweg, Berlin, Heidelberg.
- Skulmowski, A., Bunge, A., Kaspar, K., and Pipa, G. (2014). Forced-choice decision-making in modified trolley dilemma situations: a virtual reality and eye tracking study. *Frontiers in Behavioral Neuroscience*, 8, 426.