Proceedings of the Society of Automotive Engineers World Congress 2016. April 12-14, 2016. Detroit, MI

From Trolley to Autonomous Vehicle: Perceptions of Responsibility and Moral Norms in Traffic Accidents with Self-Driving Cars

Li, J. Stanford University
Cho, M. J. Stanford University
Zhao, X. Brown University
Ju, W. Stanford University
Malle, B. F. Brown University

Abstract

Autonomous vehicles represent a new class of transportation that may be qualitatively different from existing cars. Two online experiments assessed lay perceptions of moral norms and responsibility for traffic accidents involving autonomous vehicles. In Experiment 1, 120 US adults read a narrative describing a traffic incident between a pedestrian and a motorist. In different experimental conditions, the pedestrian, the motorist, or both parties were at fault. Participants assigned less responsibility to a self-driving car that was at fault than to a human driver who was at fault. Participants confronted with a self-driving car at fault allocated greater responsibility to the manufacturer and the government than participants who were confronted with a human driver at fault did. In Experiment 2, 120 US adults read a narrative describing a moral dilemma in which a human driver or a self-driving car must decide between either allowing five pedestrians to die or taking action to hit a single pedestrian in order to save the five. The "utilitarian" decision to hit the single pedestrian was considered the moral norm for both a self-driving and a human-driven car. Moreover, participants assigned the obligation of setting moral norms for self-driving cars to ethics researchers and to car manufacturers. This research reveals patterns of public perception of autonomous cars and may aid lawmakers and car manufacturers in designing such cars.

Introduction

Suppose that, sometime in the next decade, you see a car driving along a city street. It approaches an intersection and a pedestrian who is crossing the road is hit by the car and injured. You call the local authorities and approach the scene of the accident. Upon looking into the vehicle, you see that there is no driver in the car – instead, it is a car that drives by itself. Is the pedestrian, the manufacturer, the owner, the government or no one to blame for the traffic accident?

Autonomous or "self-driving" vehicles face not only technological challenges but social ones as well. One major concern surrounds the perception of traffic incidents, as illustrated by the preceding scenario. Most people don't have an immediate answer to such questions when they are posed. Autonomous vehicles have the potential to reduce the number of automobile accidents that occur on our roads; nevertheless, traffic accidents with autonomous cars will occur. Google reports their self-driving car has had one minor accident attributed to human error every 250,000 km since 2009 [1]. In comparison, the US government reports 185 crashes every 100 million miles, or 0.287 crashes every 250,000 km in 2009, which includes accidents of high severity [2]. How the public will respond to traffic incidents that occur with autonomous vehicles influences their adoption and development (e.g., [3,4]). Factors discouraging adoption of a new technology often outweigh factors encouraging adoption. In early studies of PC users, for example, social pressure and restrictive regulations motivated many potential users to not purchase, even in the presence of reasons to purchase [5]. In addition, the U.S. liability system punishes "sins of commission" rather than "sins of omission." ([6]; cf. [7]) That is, the negative effects of introducing a new technology (e.g., a death resulting from a new airbag) are punished more severely than the negative effects of failing to introduce a new technology (e.g., a death that could have been prevented by a new airbag that was not released). Therefore, one factor discouraging adoption of self-driving cars is public uncertainty regarding the liability of accidents with this type of car.

Another potentially powerful factor discouraging adoption of self-driving cars is people's uncertainty and discomfort regarding machines that might make decisions that harm human beings. Such decisions may very well be required of self-driving cars in rare but disconcerting traffic situations: when a more severe damage can be averted only by actively causing a less severe damage. Often cast as "moral dilemmas" [8, 9] when a person is the decision maker, these situations can reveal extant patterns of moral norms and responsibility judgments vis-à-vis autonomous cars. We undertake a first step toward identifying such patterns in this paper, hoping to inform lawmakers and manufacturers in designing mechanisms and regulations for self-driving cars.

Who is Responsible?

When examining how people assign moral and legal responsibility to autonomous cars for accident damage, we can distinguish between at least three targets of responsibility (Table 1). First, responsibility for accident damage could be ascribed to parties that produced the technology in question. Car manufacturer may be held responsible because of an "ultimate responsibility for their product." [10] Programmers, as part of the "producer" broadly construed, make decisions about an automated vehicle's operation ahead of time [11], and their designs take into account regulations valid at the time of design. Second, responsibility could be allocated to the human owner(s) in light of the risks they accept through using the autonomous vehicle, regardless of an ability to intervene [10]. Third, responsibility could be perceived to reside in the autonomous car itself if its decisions and actions

are treated similarly to those of a human agent. People tend to perceive agency and human-like attributes in a wide range of lifeless objects (e.g., computers and moving shapes) without immediately directing their attention to the designer of those objects [12, 13]. It is unclear whether people would also ascribe agency (and potentially the ability to make important decisions) to autonomous vehicles.

Research Question: To which target will people allocate responsibility for an accident involving an autonomous car—to the producer, the owner, or the car itself?

Table 1. Potential responsibility targets for an autonomous vehicle accident.

Target for Responsibility	Description	Potential Basis for Liability
Car manufacturer, government	Liability based on technology introduction	Commission
Car owner, user	Liability based on technology use or based on failure to intervene	Ownership
Car itself	Liability based on technology being an independent actor capable of moral decision-making	Agency
No one	No liability is assigned	N/A

Do Moral Norms Apply to Cars?

Moral issues with autonomous vehicles and robots are gaining attention as the application of these technologies grows. Cars may be considered to be moral agents based on their domain of operation rather than their ability to have beliefs, intents and judgments. In the immediate future, the behavior of an autonomous car may be engineered as lane-keeping, distance-following and obstacle-avoidance algorithms. But autonomous vehicles may be developed with high levels of decision-making ability [14]. Higher level modules that control selective activation and deactivation of lower level modules have been designed to meet the control requirements of an autonomous vehicle capable of handling diverse terrain such as road traffic [15]. Whether such capacities include explicit moral decision-making is as yet unclear because it is not an easy feat to code morals into machines [16, 17]. Nevertheless, scholars in the field of road vehicle automation have asserted that the domain of road vehicle travel involves risk and that the allocation of such risk constitutes an ethical decision [11]. Clearly, any current or near-future autonomous system lacks the wealth of knowledge people use to make moral decisions, such as characteristics of victims or subtle features of the environment. However, a car's actions in road traffic can have moral consequences, and if an algorithm or artificial intelligence faces a situation with moral consequences, it is possible that its actions would be considered motivated by morals [16]. This effect may be particularly salient with autonomous vehicles because people are more likely to attribute intentional agency to a moving object when they lack control over its movement [18].

If people treat an autonomous car as an agent in this way, they may apply familiar moral norms to this agent, just as they respond to robots that cheat, lie or perform morally questionable actions as they would to humans (e.g., [19]). Alternatively, people could apply different norms to an autonomous car. For example, people more strongly expect a robot, compared to a human, to solve a moral dilemma situation by making a utilitarian decision—sacrificing one person while saving four [20].

Research Question: Do people expect autonomous vehicles to take utilitarian actions when facing moral dilemmas?

Experimental Paradigm

Two online text-based experiments assessed public perception of responsibility (Study 1) and moral norms (Study 2) in accidents involving autonomous vehicles. Text-based methodologies have been employed in the domain of traffic accident analysis to assess public perception of safety issues such as running red lights [21], speeding [22] and seat belt usage [23]; and they have been extensively employed in research on human moral judgment (e.g., [24]).

Study 1

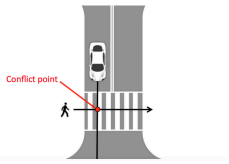
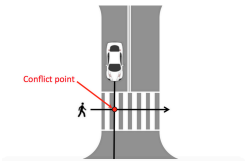
Method

One hundred and twenty participants (60 M, 60 F) between 19 and 64 years old (M = 32.2; SD = 9.8) were recruited from Amazon Mechanical Turk (MTurk) for an online experiment. Participants with IP addresses originating in the US were selected. Participants had between 0 and 48 years of experience driving (M = 14.4, SD = 10.3). The highest level of education (completed or in progress) was college or a Bachelor's degree for 86 participants, high school for 26 participants and post-graduate for 7 participants. Participants were paid $.80 ($16/hr) for the study that took approximately 3 minutes.

Two MTurk online tasks (one for females, one for males) were created to achieve gender balance in the study. Each task directed participants to a common Qualtrics survey, described as being about "traffic scenarios" so as to be applicable to both autonomous and normal cars. Participants were randomly assigned to read one of six scenario descriptions as part of a 2 (Motorist type: Autonomous car or Normal car) × 3 (Fault: Motorist-fault, Pedestrian-fault or Both-fault) between-participants study with separate randomization pools for females and males. A full factorial design was used to avoid the effects of presentation order (cf. [25]). Participants then answered questions about responsibility attribution in an automobile incident ("How much are each of the following parties to blame for the traffic accident?"; 10-point Likert scale from "No Blame" to "Full Blame") with the following judgment targets: the pedestrian, the manufacturer of the car, the government agency that regulates traffic standards in the area, the developer of the software algorithm for the car, the owner of the car, and the car itself. The last three targets were omitted in the human-

driver condition and replaced by the driver of the car. The manufacturer and software developer were both included as judgment targets, although in the analysis phase we test whether they both constitute the "producer" of the autonomous car. The presentation order of the choices was randomized with the last choice always being "Other (please specify)." Participants were also asked, in an open-ended text field, which party they felt was most to blame for the traffic incident and why.

Table 2: List of vignettes for Study 1 (variations in experimental manipulations are highlighted in blue for Fault and in red for Motorist Type).

Party-at-fault	Autonomy Fully autonomous car [a]	Human-driven car
Text: Pedestrian-fault	An autonomous car is driving by itself along a city street to pick up its owner. It approaches an intersection. A pedestrian does not notice the car because they are not paying attention. The pedestrian crosses the intersection when they are not supposed to. The pedestrian is hit by the car and injured.	A person is driving along a city street. Their car approaches an intersection. A pedestrian does not notice the car because they are not paying attention. The pedestrian crosses the intersection when they are not supposed to. The pedestrian is hit by the car and injured.
Text: Motorist-fault	An autonomous car is driving by itself along a city street to pick up its owner. It approaches an intersection. The car's object identification system does not detect a pedestrian because the pedestrian is not distinguishable from the background. The car enters the intersection when it is not supposed to. The pedestrian is hit by the car and injured.	A person is driving along a city street. Their car approaches an intersection. The driver does not notice a pedestrian who is crossing because they are not paying attention. The driver crosses the intersection when they are not supposed to. The pedestrian is hit by the car and injured.
Text: Both-fault	An autonomous car is driving by itself along a city street to pick up its owner. It approaches an intersection. The car does not detect a pedestrian who is crossing because the pedestrian is not distinguishable from the background. The pedestrian does not notice the car because they are not paying attention. Both the car and the pedestrian enter the intersection when they are not supposed to. The pedestrian is hit by the car and injured.	A person is driving along a city street. Their car approaches an intersection. The driver does not notice the pedestrian and the pedestrian does not notice the car because both are not paying attention. Both cross the intersection when they are not supposed to. The pedestrian is hit by the car and injured.
Image: All conditions		

[a] Scenarios for a fully autonomous car were preceded by "For the following scenario, please imagine that this is in the future and it is legal for autonomous cars to drive themselves." Images were created by a member of the research team, incorporating a car image © Jibran. Available: http://www.clker.com/clipart-white-car-top-view.html

Materials

In all experimental conditions, the text description of the traffic scenario was accompanied by an illustration (Table 2). The text varied for motorist type by referring to a fully autonomous car in the Autonomous car condition or a human driver in the Normal car condition. The text also varied by fault condition, following [26]: Pedestrian-fault, Motorist-fault, or Both-fault. The critically varying phrases were "they are not paying attention" (pedestrian or driver) and "the car does not detect a pedestrian" (autonomous vehicle). Pedestrian detection is a key feature of autonomous vehicles with some systems performing very accurately (e.g., [27]), while it is still conceivable that such systems may fail due to the high variability of contexts in which the system needs to function [28]. Human failure to pay attention was used instead of a description of a specific traffic situation (such as running a red light) to maintain consistent scenarios across vignettes and because distracted driving is one of the top risk factors for traffic accidents [29-31]. Several potentially varying factors were held constant (but would be interesting variables in future research): the illustration, accident severity ("seriously injured"), information about why the autonomous car was at fault and how the car operates, location (local street), and precipitating situation (pedestrian crossing the street). The last two variables were most prevalent in studies of real-life pedestrian-motor vehicle crashes [26].

Results

Manipulation check

Planned simple contrasts conducted in R showed that, as intended, greater responsibility was assigned to the pedestrian in the Pedestrian-fault condition than in the Motorist-fault condition for the Normal car, $t(57) = 5.93$, $p < .0001$, $d = 1.6$. Greater responsibility was also assigned to the pedestrian in the Pedestrian-fault condition than in the Both-fault condition for the Normal car, $t(57) = 3.14$, $p = .003$, $d = .83$. Similarly, greater responsibility was assigned to the driver in the Motorist-fault condition than in the Pedestrian-fault condition for the Normal car, $t(56) = 6.04$, $p < .0001$, $d = 1.6$. No difference was found in responsibility assigned to the driver between Motorist-fault and Both-fault conditions, $t(56) = 0.91$, $p = .369$.

Responsibility for accident

Overall, an autonomous car was judged to be less responsible than a human driver in a traffic accident. A linear regression was performed on rating of driver responsibility with Motorist type, Fault and their interaction as predictors, treating the condition of Pedestrian-fault and Normal car as the reference category, $F(5,113) = 9.58$, $p < .0001$, $d = 1.8$. In the Motorist-fault condition, people assigned less blame to an Autonomous car at fault compared to the driver of a Normal car at fault, while there was no difference in blame for the Pedestrian-fault condition, interaction effect $t(113) = -3.145$, $p = .002$, $d = .59$ (Figure 1).

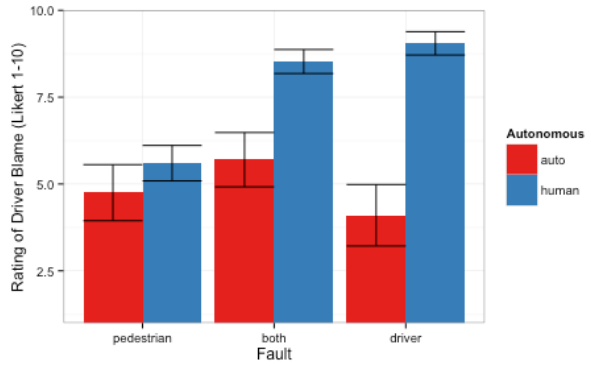


Figure 1. Participant ratings of motorist blame for a traffic accident with an Autonomous or Normal car.

Targets of responsibility

Motorist type influenced the selection of who was responsible for the accident. Multiple linear regression analyses were performed with Motorist type, Fault and their interaction as predictors and responsibility assignments as the response variables. Lower responsibility was assigned to the owner of an Autonomous car than to the owner of a Normal car, $t(113) = -2.98$, $p = .004$, $d = .56$. Greater responsibility was assigned to the manufacturer/software developer with an Autonomous car than with a Normal car, $t(114) = 6.11$, $p < .0001$, $d = 1.14$ (Figure 2), as well as greater responsibility to the government, $t(114) = 3.37$, $p = .001$, $d = .63$. As expected, ratings for the manufacturer and software developer were found to be similar and therefore their items were combined, Pearson's $r = 0.68$, $t(58) = 7.144$, $p < .001$. No interaction effects were found: participants assigned greater responsibility to the government and the manufacturer with an Autonomous car than with a Normal car even when the car was not at fault.

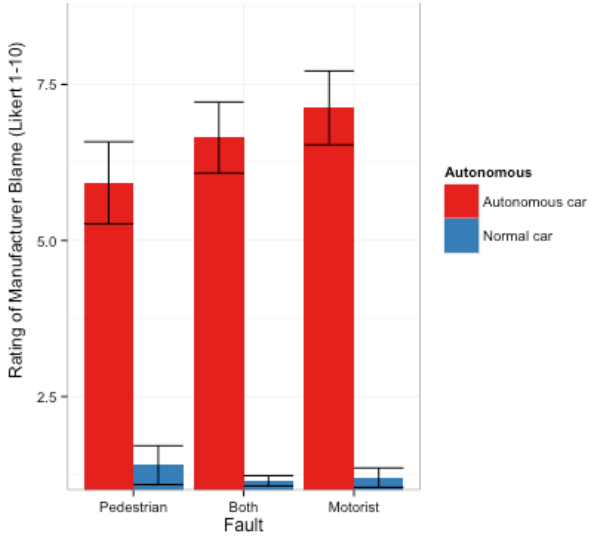


Figure 2. Participant ratings of manufacturer blame for a traffic accident with an Autonomous or Normal car.

Difficulty

Motorist type influenced judgment difficulty in the one critical condition. Responsibility was more difficult to assign with the Autonomous car compared to the Normal car for the Motorist-fault condition, $t(112) = 2.19$, $p = .031$, $d = .41$ but not for other conditions.

Qualitative responses

With a Normal car, people's unconstrained responses to the question of which target has the greatest responsibility in the accident were as expected: the pedestrian was mentioned as most responsible in the Pedestrian-fault condition and the driver was mentioned as most responsible in the Motorist-fault condition. In contrast, judgments varied for an Autonomous car: both pedestrian and driver were mentioned as most responsible in the Pedestrian-fault condition while various parties were mentioned in the Motorist-fault and Both-fault conditions. Table 3 lists examples.

Table 3. Sample responses for party most responsible for accident.

Motorist	Fault	Sample response
Normal car	Pedestrian-fault	"The pedestrian. He or she should be watching where he is walking! Always look both ways and all that"
	Motorist-fault	"The driver of the car is fully at blame because he wasn't paying attention. He should have stopped at the crosswalk to let the pedestrian pass"
	Both-fault	"The driver, because they are operating the car and should be aware of their surroundings more than a pedestrian"
Autonomous car	Pedestrian-fault	"The pedestrian is most to blame because they were not paying attention and crossed when they shouldn't have"
	Motorist-fault	"The designer of the automated car should be held responsible because his object identification system isn't good enough yet"
	Both fault	"The car and related companies should take most (70/30) of the blame. The pedestrian has to assume some responsibility for crossing the road without paying attention."

Discussion

Study 1 demonstrated that people did not hold an autonomous car as responsible for an accident as they held a human driver in the same situation. In particular, even when the autonomous car was at fault, people blamed it much less than they blamed the human driver when at fault. However, when the pedestrian was at fault, blame for both autonomous car and human driver was low. In addition, people assigned greater responsibility to the manufacturer and to the government for a traffic incident with a self-driving car compared to a human-driven car. This pattern held regardless of whether the motorist or the pedestrian was at fault.

People therefore appear reluctant to consider an autonomous car as a viable agent that can be blamed for an accident; instead, they step back and identify producers of this car (manufacturers and programmers) and the government as the ones to blame. People may still consider autonomous vehicles as potentially dangerous new technology and therefore put responsibility into the hands of those who can (now) prevent or (in the future) correct any damage this technology causes. These exploratory results should be taken as tentative because blame assigned to all parties was in the middle of the scale in the autonomous car conditions; thus, respondents may have been generally uncertain about how to assign responsibility in the study. This does not make the finding insignificant, however, because if an accident like this happened right now, the general public (and possible jury members) would display similar uncertainty.

In the current study, blame ratings for each party were analyzed independently from one another, even though responsibility and blame are generally allocated across multiple parties depending on their contributions to the outcome [32]. By presenting two more possible targets of responsibility in the autonomous versus human driver conditions, we may have prompted people to distribute a "fixed pool" of blame more sparsely across the parties in the autonomous condition. However, the mean blame was 5.66 in the autonomous car condition and 4.20 in the human condition, making sparse distribution of blame very unlikely.

One limitation of the present study is that only one element was varied in the manipulation of fault. Specific causes of product failure influence consumer reaction [33] so it is possible that different causes of autonomous vehicle fault (e.g., algorithm limitation, electrical malfunction) could influence responsibility attribution.

A second factor that could hinder adoption of autonomous vehicles is the uncertainty surrounding their behavior in moral dilemmas. Study 2 investigated how moral norms apply to cars that incorporate algorithms for high-level decision-making.

Study 2

Method

One hundred and twenty participants (60 M, 60 F) between 19 and 64 years old ($M = 33$; $SD = 10.4$) were recruited from Amazon Mechanical Turk (MTurk) in the same manner as in Study 1. Participants had between 0 and 48 years of experience driving ($M = 14.4$, $SD = 10.3$). The highest level of education completed or in progress was college or a Bachelor's degree for 93 participants, high school for 21 participants and post-graduate for 6 participants. Participants were paid $.80 ($16/hr) for the study that took approximately 3 minutes. Studies 1 and 2 were run concurrently but employed the same randomization pool so that no participant who completed one study could complete the other.

The procedure was identical to Study 1, except with different manipulations and measures. Participants were randomly assigned to read one of six vignettes as part of a 3 (Motorist type: Autonomous car, Handover or Normal car) × 2 (Victim: Single pedestrian or Motorist) online between-participants study. Study 2 added a "Handover" level for Motorist type: participants in this condition read a vignette in which the human driver takes over control of the autonomous car at the time of action. Victim referred to whether the vignette gave the motorist the choice of harming either a single pedestrian or the motorist to save five pedestrians. After reading the description of a moral dilemma scenario, participants answered a question on moral norms ("Should the [car / person] steer to the lane with [one person / the large truck] or stay in the lane with five people?") and one question on the appropriate target of responsibility for deciding an autonomous car's actions ("Who should determine the best way for [an autonomous car / a person] to respond to this situation?"). The presentation order of judgment targets was randomized.

Materials

Text descriptions with an accompanying illustration (Appendix A) were used as in Study 1. We used the classic "trolley problem" [9] as a starting point but made two major modifications. First, we changed the setting to be a traffic scenario that could realistically involve autonomous vehicles. Second, whereas the classic trolley problem describes the scene from the perspective of the main actor (such that the reader is placed "in their shoes"), we described the scene in the third person to assess general moral norms rather than individual preferences for action.

Results

Moral norms

Across conditions, three quarters (75%) of participants said the motorist should switch lanes, 7% said it should stay in the current lane, 11% said either was acceptable and 8% said they didn't know. To examine whether these percentages depended on car type, a logistic regression analysis was conducted in R with Motorist type, Victim and their interaction as predictor variables. The percentage of participants who selected the (utilitarian) choice of switching lanes did not significantly differ by Motorist type or Victim for either the omnibus test, $\chi^2(2, N = 120) = 6.14$, $p = .73$, or main effect tests. 82.5% of participants who read a situation with an Autonomous car said the car should make a switch; 75% in the Handover condition said the person should make the switch; and 67.5% who read the Normal car condition said the person should make the switch. Although not significant, these percentages and their relative order are similar to Malle et al. [19] (78% of participants said a robot should select a utilitarian choice and 65% said a human should). Systems with greater automation may be more strongly expected to make utilitarian decisions than those with more human control.

Table 4: Frequencies of selecting options for action in a moral dilemma.

	Experimental conditions					
	Normal car		Handover		Autonomous car	
Moral norm	Single pedestrian victim	Motorist victim	Single pedestrian victim	Motorist victim	Single pedestrian victim	Motorist victim (car)
Non-utilitarian (injure 5 people in current lane)	0	2 (10%)	0	3 (15%)	1 (5%)	2 (10%)
Utilitarian (injure victim)	13 (65%)	14 (70%)	16 (80%)	14 (70%)	17 (85%)	16 (80%)
Either	3 (15%)	3 (15%)	2 (10%)	3 (15%)	1 (5%)	1 (5%)
Don't Know	4 (20%)	1 (5%)	2 (10%)	0	1 (5%)	1 (5%)

Setting moral norms

To examine people's judgments of who should determine moral norms for autonomous vehicles, we used multiple linear regression analyses with Motorist type and Victim as a predictor variables and responsibility assignments as the response variables. No differences were found in the responsibility assigned to the public or government (Table 5, rows 1-2). Greater responsibility was assigned to ethics researchers for deciding the moral behavior of an Autonomous car compared to a Normal car, $t(114) = 3.48$, $p < .001$, $d = .65$. Similar main effects were found for manufacturer and software developers. The manufacturer and the software developer were judged as more responsible for deciding moral norms for an Autonomous car than for a Normal car, $t(114) = 1.94$, $p = .055$, $d = .36$ and $t(114) = 3.28$, $p < .01$, $d = .61$. Conversely, the car owner was judged as less responsible for deciding moral norms for an Autonomous car or a Handover than they were for a Normal car, $t(114) = -2.68$, $p < .01$, $d = -0.50$ and $t(114) = -2.91$, $p < .01$, $d = -0.55$. The car driver was also judged as less responsible for deciding moral norms for an Autonomous car or a Handover than for a Normal car, $t(114) = -3.25$, $p = .002$, $d = -0.61$ and $t(114) = -2.30$, $p = .023$, $d = -0.43$.

Table 5: Ratings of responsibility for an ethical decision, means (*SD*). Coded - 0: "Definitely No", 1: "Maybe", 2: "Definitely Yes". Highest mean ratings for each condition are in bold.

	Human-driven	Autonomous	Handover
Public	0.575 (.636)	0.700 (.687)	0.475 (.554)
Government	0.500 (.641)	0.575 (.675)	0.350 (.483)
Ethics Researchers	0.675 (.616)	**1.28 (.679)**	0.900 (.591)
Car Manufacturer	0.625 (.740)	0.875 (.723)	0.375 (.490)
Software Developer	0.725 (.816)	1.175 (.747)	0.825 (.636)
Car Owner	1.225 (.698)	0.850 (.770)	0.900 (.672)
Car Driver	**1.525 (.679)**	1.05 (.876)	**1.45 (.639)**

Interaction effects were also found for manufacturer and software developer situations. Greater responsibility was assigned to the manufacturer when the victim was a Motorist compared to a Single pedestrian for a Normal car, while no difference in responsibility was found with a Handover, $t(114) = -2.05$, $p = .042$, $d = -0.38$. This same interaction was found for the software developer, $t(114) = -2.49$, $p = .014$, $d = -0.47$. The scenario where the victim was the Motorist involved the car hitting a large truck and may have prompted participants to consider the car's structural strength.

Discussion

In Study 2, it was more acceptable for an autonomous car to make a utilitarian decision to injure either a single pedestrian or a motorist than it was to decide to injure five pedestrians. In addition, autonomous cars were expected to make utilitarian decisions more so than people were, though this effect did not reach significance. It is possible that an autonomous vehicle that is portrayed as having more agent-like features (e.g., a voice, name, perhaps even a virtual face on a screen) would be more strongly expected to follow moral utilitarian norms than the car described in this study. The two options used for the dilemma (to "stay in the lane and injure one person" or "switch lanes to injure one instead of five people") were also selected because of their similarity to those used in past work, but may have favored the switching action too heavily, since almost all participants selected that option. Wording effects may also have influenced results, particularly as terms such as "should", "wrong" and "permissible" activate separate psychological constructs [24].

Conclusion

Two experiments evaluated public perceptions of accidents involving autonomous vehicles. In Experiment 1, people ascribed less responsibility to an autonomous car when at fault than to a human driver when at fault. People did not treat the autonomous car as an independent moral agent responsible for its own actions. They instead identified the car manufacturer and the government as the parties responsible for the technology. In Experiment 2, ethics

researchers and car manufacturers were judged as having a greater obligation for deciding moral norms governing the actions of a self-driving car compared to a human driver. Moreover, there was a trend that people preferred autonomous vehicles to make a utilitarian decision when faced with the dilemma of possibly sacrificing one life to save five others.

References

1. Hood, J., "Google Defends Its Self-Driving Cars' Accident Rate," *Consumer Affairs*, May 12, 2015. Available: http://www.consumeraffairs.com/news/google-defends-its-self-driving-cars-accident-rate-051215.html
2. Barry, K., "Safety in Numbers: Charting Traffic-Safety and Fatality Data," *Car and Driver*, May 2011. Available: http://www.caranddriver.com/features/safety-in-numbers-charting-traffic-safety-and-fatality-data
3. Lin, P., "The Ethics of Autonomous Cars," *The Atlantic,* Oct 8, 2013. Available: http://www.theatlantic.com/technology/archive/2013/10/the-ethics-of-autonomous-cars/280360/
4. Beiker, S., "Legal Aspects of Autonomous Driving," *Santa Clara Law Review, 52*(4): 1145-1156, 2012.
5. Venkatesh, A. and Brown S.,"A Longitudinal Investigation of Personal Computers in Homes: Adoption Determinants and Emerging Challenges," *MIS Q, 25*(1): 71–102, 2012.
6. Malhotra, N., Shotts, K. and Melvin, S., "The Nut Behind the Wheel' to 'Moral Machines': A Brief History of Auto Safety," Stanford Graduate School of Business Case No. ETH4, 2014.
7. Chipman, I., "Exploring the Ethics Behind Self-Driving Cars," Stanford Graduate School of Business, 2015. Available: http://www.gsb.stanford.edu/insights/exploring-ethics-behind-self-driving-cars
8. Marcus, R. B., "Moral dilemmas and consistency," *The Journal of Philosophy, 177*(3): 121-136, 1980.
9. Foot, P., "The Problem of Abortion and the Doctrine of the Double Effect," *Oxford Review, 5*, 1967.
10. Hevelke, A. and Nida-Rümelin, J., "Responsibility for Crashes of Autonomous Vehicles: An Ethical Analysis,' *Sci. Eng. Ethics, 21*(3): 619-630, 2014.
11. Goodall, N., "Ethical Decision Making During Automated Vehicle Crashes," *Transportation Research Record: Journal of the Transportation Research Board, 2424*: 58-65, 2014.
12. Złotowski, J., Proudfoot, D., Yogeeswaran, K. and Bartneck, C., "Anthropomorphism: opportunities and challenges in human–robot interaction," *International Journal of Social Robotics, 7*(3): 347-360, 2015.
13. Reeves, B. and Nass, C., *The Media Equation*, Cambridge: Cambridge University Press, 1996.
14. Parasuruman, R., Sheridan, T. and Wickens, C., "A Model for Types and Levels of Human Interaction with Automation," *IEEE Transactions of Systems, Man and Cybernetics - Part A, 30*(3): 2000.
15. Payton, D., "An architecture for reflexive autonomous vehicle control," In *Robotics and Automation. Proceedings. 1986 IEEE International Conference on*, 3: 1838-1845, 1986. IEEE.
16. Coeckelbergh, M., "Moral Appearances: Emotions, Robots, and Human Morality," *Ethics and Information Technology*, *12*(3): 235-241, 2010.
17. Malle, B. F. and Scheutz, M., "Moral competence in social robots," In *IEEE International Symposium on Ethics in Engineering, Science, and Technology*, 30–35, 2014. Chicago, IL: IEEE.
18. Barrett, J. and Johnson, A., "The Role of Control in Attributing Intentional Agency to Inanimate Objects," *Journal of Cognition and Culture, 3*(3): 208-217, 2003.
19. Breazeal, C., "Toward Sociable Robots," *Robotics and Autonomous Systems, 42*(3): 167–175, 2003.
20. Malle, B., Scheutz, M., Arnold, T., Voiklis, J. et al., "Sacrifice One for the Good of Many? People Apply Different Moral Norms to Human and Robot Agents," *HRI '15*, 117-124, 2015.
21. Public Opinion Survey Traffic and Public Safety Lafayette Parish, 2001. Available: http://www.lafayettela.gov/TrafficAndTransportation/SafeLight/SiteAssets/Files/WalkerAssociatedSurvey101101.pdf
22. Streff, F. and Kostynuik, L., "Survey of Public Perception of Traffic Law Enforcement in Michigan," Transportation Research Institute UMTRI-2000-37, 2000. Available: http://deepblue.lib.umich.edu/handle/2027.42/1338
23. Stern, D., *A Survey to Determine the Influence of Traffic Accident Severity and Selected Variables on Seat Belt Usage by the General Public*, 1977.
24. Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M. and Cohen, J. D., "An fMRI investigation of emotional engagement in moral judgment," *Science*, *293*(5537): 2105–2108, 2001.
25. O'Hara, R.,"Wording Effects in Moral Judgment," *Judgment and Decision Making, 5*(7): 547-554, 2010.
26. Ulfarsson, G., Kim, S. and Booth, K., "Analyzing fault in pedestrian–motor vehicle crashes in North Carolina," *Accident Analysis and Prevention, 42*(6): 1805-1813, 2010.
27. Broggi, A., Bertozzi, M., Fascioli, A. and Sechi, M., "Shape-Based Pedestrian Detection," *Proceedings of the IEEE Intelligent Vehicles Symposium*, 215-220, 2000.
28. Gavrila, D., "Pedestrian Detection from a Moving Vehicle," *ECCV*, 2000.
29. Young, K. and Salmon, P., "Sharing the Responsibility for Driver Distraction Across Road Transport Systems: A Systems Approach to the Management of Distracted Driving," *Accident Analysis & Prevention, 74*, 350-359, 2015.
30. Stewart, A., "Attributions of Responsibility for Motor Vehicle Crashes," *Accident Analysis & Prevention, 37*(4): 681-688, 2005.
31. Atchley, P., Hadlock, C. and Lane, S., "Stuck in the 70s: The Role of Social Norms in Distracted Driving," *Accident Analysis & Prevention, 48*: 279-284, 2012.
32. Gerstenberg, T. and Lagnado, D. A., "Spreading the Blame: The Allocation of Responsibility Amongst Multiple Ag ents," *Cognition, 115*(1): 166-171, 2010.
33. Folkes, V., "Consumer Reactions to Product Failure: An Attributional Approach," *Journal of Consumer Research, 10*(4): 398-409, 1984.

Contact Information

Jamy Li, jamy@stanford.edu

Appendix

Appendix A: Vignettes used in Experiment 2 (variations in experimental manipulations are highlighted in blue for Switch Target and in red for Autonomy)

Vignette component	Motorist type	Victim: Single pedestrian	Victim: Motorist
Text	Fully autonomous car[b]	An autonomous car that is driving to pick up its owner is traveling along a road with two lanes when a tire blows out. The car brakes but detects that it won't be able to stop before hitting five people who are crossing the road in the car's current lane. The car can steer to move itself from one lane to the other lane where it would hit one person who is crossing at this moment. Whether the car stays in its current lane or moves to the other lane, any pedestrian the car will hit would be seriously injured.	An autonomous car that is driving to pick up its owner is traveling along a road with two lanes when a tire blows out. The car brakes but detects that it won't be able to stop before hitting five people who are crossing the road in the car's current lane. The car can steer to move itself from one lane to the other lane where it would hit a large truck that is driving towards it at this moment. If the car stays in its current lane, the pedestrians the car will hit would be seriously injured. If the car moves to the other lane, the car would be seriously damaged in the collision with the truck.
	Handover action[c]	An autonomous car that is driving by itself is traveling along a road with two lanes when a tire blows out. The person sitting in the driver's seat of the car takes over and brakes but realizes that the car won't be able to stop before hitting five people who are crossing the road in the car's current lane. The person in the car can manually steer to move the car from one lane to the other lane where it would hit one person who is crossing at this moment. Whether the car stays in its current lane or moves to the other lane, any pedestrian the car will hit would be seriously injured.	An autonomous car is driving by itself along a road with two lanes when a tire blows out. The person sitting in the driver's seat of the car takes over and brakes but realizes that the car won't be able to stop before hitting five people who are crossing the road in the car's current lane. The person in the car can manually steer to move the car from one lane to the other lane where it would hit a large truck that is driving towards it at this moment. If the car stays in its current lane, the pedestrians the car will hit would be seriously injured. If the car moves to the other lane, the driver of the car would be seriously injured in the collision with the truck.
	Normal car	A car is traveling along a road with two lanes when a tire blows out. The driver brakes but realizes that the car won't be able to stop before hitting five people who are crossing the road in the car's current lane. The driver can steer to move the car from one lane to the other lane where it would hit one person who is crossing at this moment. Whether the car stays in its current lane or moves to the other lane, any pedestrian the car will hit would be seriously injured.	A car is traveling along a road with two lanes when a tire blows out. The driver brakes but realizes that the car won't be able to stop before hitting five people who are crossing the road in the car's current lane. The driver can steer to move the car from one lane to the other lane where it would hit a large truck that is driving towards it at this moment. If the car stays in its current lane, the pedestrians the car will hit would be seriously injured. If the car moves to the other lane, the driver of the car would be seriously injured in the collision with the truck.
Image	All motorist types