Vehicle Automation and the Duty to Act

Noah J. Goodall

Presented at the 21st World Congress on Intelligent Transport Systems. Detroit, MI, September, 2014.
Preprint copy, formatted for Web

Abstract
The act of driving always carries some level of risk. With the introduction of vehicle automation, it is probable that computer-driven vehicles will assess this changing level of risk while driving, and make decisions as to the allowable risk for itself and other road users. In certain situations, an automated vehicle may be forced to select whether to expose itself and its passengers to a small risk in order to protect other road users from an equal or greater amount of cumulative risk. In legal literature, this is known as the duty to act. The moral and legal responsibilities of an automated vehicle to act on the behalf of other road users are explored.

INTRODUCTION

Road vehicle automation has progressed rapidly in recent decades. Many technologies designed to assist humans by automating aspects of the driving process are in today's vehicles, including anti-lock brakes, cruise control, stability control, and lane keeping. Today's prototype vehicles employ 360 degree sensing *(1)* and utilize probabilistic models to anticipate the movements of nearby vehicles and plan accordingly *(2)*. The constant monitoring of the environment and quicker reaction times of computerized control are expected to significantly reduce the estimated 93% of crashes that are the result of human error *(3, 4)*.

Several automakers insist that the human driver will ultimately be responsible for monitoring the vehicle and the roadway, and in the event of a crash, the driver will be available to take control with little notice. This is the scenario described by the National Highway Safety Administration for level 2 automation, with levels 3 and 4 allowing the driver significant warning before returning to the driving task *(5)*. Yet recent research suggests that it is unrealistic to assume that a human passenger can effectively monitor an automated vehicle's. On test tracks, passengers in automated vehicles spend more time looking away from the road than those in traditional vehicles *(6)*.

In the absence of monitoring by its human driver, an automated vehicle could attempt to predict dangerous situations and alert the driver in advance. This is difficult to implement in practice, as humans may require substantial warning time. Drivers in simulators exhibit excessive swerving, braking, and failure to check blind spots when given seven seconds of warning to respond to a stopped vehicle after riding in an automated vehicle *(7)*. Stanford's Sven Beiker has mentioned studies where blindfolded drivers in automated vehicles required five to six seconds to assess the situation when the blindfolds were removed, and up to a minute before their driving matched an un-blindfolded control subject *(8)*. No algorithm can

Noah J. Goodall, Research Scientist, Virginia Center for Transportation Innovation and Research
530 Edgemont Road, Charlottesville, VA 22903, noah.goodall@vdot.virginia.gov

possibly predict with certainty the movements of other vehicles, pedestrians, cyclists, or wildlife six seconds into the future. In many potentially dangerous situations, the computer must take evasive action without waiting on the approval of its human driver.

In regular operations, highly-automated vehicles use path planning algorithms to develop and select from a set of candidate paths *(1)*. A similar strategy can be used in near-crash situations to determine the best way to avoid collisions. In most situations, selecting the safest path should be straightforward, as the best path is one with an exceptionally low probability of collision. In situations with a great deal of uncertainty or when a crash is unavoidable, the decision may become much more complex. Given a choice between emergency braking to avoid a deer in the road, techniques from risk analysis can be used to balance the uncertainty of a choice (e.g. the deer may move on its own, or the following truck driver may not notice the sudden braking in time to stop) with the severity of its possible outcomes. When a crash is unavoidable, crash severity models could be used to determine the safest alternative. This assumes that the goal is to minimize the cumulative injury, expressed quantitatively. Using this metric exclusively leads to some morally ambiguous results. One example that has been provided is that when given a choice between colliding with two different vehicles, an automated vehicle may consistently choose the SUV with which it has a better crash compatibility *(9)* rather than the sedan, or worse may consistently collide with the "safer" helmeted motorcyclist over the non-helmeted rider *(10)*. These gray areas of decision making in vehicle automation can be addressed using the language of morals and ethics. Previous work has discussed the need for a moral component in an automated vehicle's path planning algorithms *(11, 12)*, as well as the challenges of encoding a moral system in software *(10)*.

An especially difficult problem concerns when an automated vehicle has the opportunity to render aid, by subjecting itself to a small risk in order to protect another roadway user. This can occur either by taking some action or, as in the example discussed in the next section, by refraining from an action. For example, should an automated vehicle, without the permission of its owner, intentionally collide at low-speed and with miniscule risk of injury in order to protect a more vulnerable pedestrian? An advanced automated vehicle will likely be aware of its environment and continuously calculating risk, and may be able to anticipate these types of situations more often than humans. Because it is software-based, its actions in these rare but important situations can be dictated ahead of time, with careful thought and consideration.

In the following sections, the obligation of an automated vehicle to the safety of other road users, referred to in legal literature as the affirmative duty to act, is examined from legal and moral perspectives.

THE MORAL DUTY TO ACT

In a recent magazine article, a vehicle automation engineer describes a crash where a distracted driver failed to notice stopped traffic ahead and collided with a stopped vehicle. The author of the article concludes that if the distracted driver's vehicle had been automated, "it would have seen the obstruction three cars ahead. It would have calculated the distance to impact, scanned the neighboring lanes, realized it was boxed in, and hit the brakes, all within a tenth of a second," thus avoiding the crash entirely *(13)*.

Changing the circumstances slightly raises a new problem. Assume that the distracted driver is driving himself in vehicle A, while the leading driver is in an automated vehicle B which is not boxed in. In the original example, B was struck at thirty miles per hour, with enough force to collide with the vehicle in front *(13)*. If B were instead automated with the ability to sense three cars ahead and behind, to calculate distance to impact, and to scan neighboring lanes, B could realize a collision was imminent and pull into an adjacent lane, avoiding the distracted driver entirely. See Figure 1 for a representation.

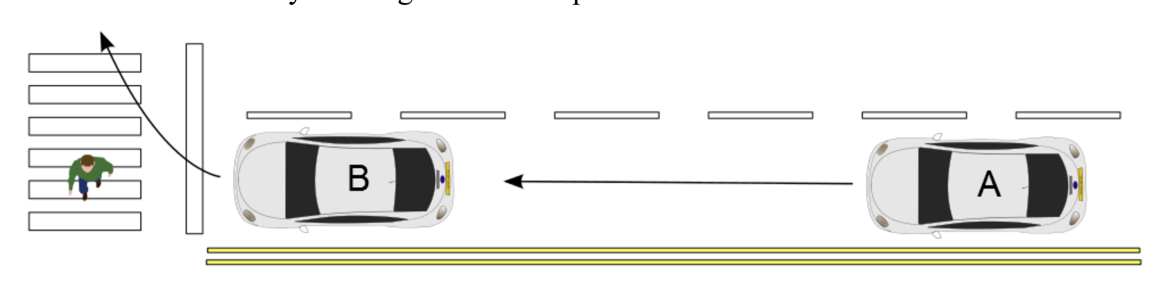

Figure 1. An impending collision between a distracted driver (A) and an automated vehicle (B).

This avoids one crash, but creates a new, possibly more dangerous crash. If the distracted driver remains distracted, he may continue through the crosswalk, striking the pedestrian. Even if the driver begins to brake, at 30 miles per hour he requires 40 feet to stop in dry conditions, and may still strike the pedestrian. Yet if the automated vehicle had remained in place, it may have absorbed much of the impact and spared the pedestrian. While a 30 mph collision with a passenger car is certainly dangerous, it is far safer within a modern vehicle than without.

This configuration creates several new problems. Should the automated vehicle be allowed to move into the adjacent lane? Does the rule change if the automated vehicle is empty? If the same situation occurred, except that the automated vehicle began in the adjacent lane but had the opportunity to block the path of the oncoming vehicle, is it obligated (or permitted) to do so? These questions are considered briefly from the perspectives of US common law and ethics.

Legal Perspective

There are many laws that may affect the example collision scenario described in the previous section, including roadway laws and product liability. The scope of this paper is limited to the affirmative duty to act under common law in the United States. For a more detailed analysis of the legality of automated vehicles, see Smith *(14)*. For a discussion of the legal issues surrounding autonomous agents generally, see Chopra and White *(15)*. Because there is no literature on the duties of automated vehicles, this section assumes that vehicles will be held to the same standards as a reasonable person in regards to the duty to act.

The affirmative duty to act refers to one's responsibilities to assist another in an emergency, and is classified as a type of negligence. In general, there is no duty to act affirmatively according to common law. First year law students are often surprised to learn that there is no duty to act unless the bystander is one of the following: contractually obligated to rescue (e.g. a life guard), in a special relationship with the victim such as a parent-child, responsible for the predicament, or subject to a special requirement such as mandatory reporting. Put simply, as long as the drowning victim is not your child, you didn't push them in, and you're not a lifeguard, you can legally walk away because you don't want to get your feet wet. There is even a disincentive to assist—one who voluntarily begins a rescue assumes the duty to continue the rescue, as they may have prevented others from initiating a second rescue. From Figure 1, Vehicle B has no duty to protect the pedestrian from harm unless they are determined to have a special relationship.

This has been acknowledged by many as a shortcoming of the legal system *(16–18)*. (For a defense of no-duty-to-act laws, see Scordato *(19)*). Obviously it's asking a great deal for a bystander to tolerate massive risk for a stranger, but to require absolutely no effort seems counter to society's preferences. Many countries in Europe and Latin America have adopted Good Samaritan laws to protect bystanders who intervene, as well as laws to enforce penalties on those who fail to render aid. Additionally, several proposals have been put forth for ways to introduce affirmative duty to act into American law *(20)*. While duty to act laws apply to persons rather than products, it may be difficult to introduce this behavior in vehicle automation without customs associated with the law.

Moral Perspective

The field of ethics provides a language with which to discuss the duty to act. This section discusses both normative and descriptive approaches.

Normative Ethics

Normative ethics describe what we as individuals and society ought to do, not what we do or prefer at any given time. While there are many relevant ethical theories regarding the duty to act, this section discusses utilitarian and deontological theory.

Utilitarianism is a form of consequentialism where an action is considered good if it produces the maximum net cumulative benefit or utility *(21)*. This strategy is popular among computer scientists due to its computability, and is thus likely to be a component of early automated vehicle ethical systems. In the collision example from Figure 1, a utilitarian vehicle would stay in its place and absorb the impact, determining that whiplash is preferable to the death of a pedestrian. In fact, a utilitarian automated vehicle would stay in place even if the impact resulted in quadriplegia for the passenger, as this would be preferable to the death of the pedestrian. This outcome would be unacceptable to most, and reveals a flaw of utilitarianism: it treats all users as equals, regardless of context. To a utilitarian, fault is irrelevant, the perspective of the moral agent is disregarded, and the sole objective is the relative utility of potential outcomes. For most people, though, context matters a great deal. The passenger in the automated vehicle may feel strongly that, since their vehicle was not at fault, then they have the right to try and avoid the collision.

Deontology is a moral theory based on adherence to a set of rules, duties, or rights *(22)*. It is often described as the opposite of consquentialism, which determines whether an action is moral based solely on the outcome rather than any underlying principle. A deontologist could argue that an automated vehicle (considered as an autonomous moral agent for simplicity) has a duty to rescue if the cost to itself (and its passengers) is comparable to the one in need of rescue. This must be balanced with a vehicle's (and its passengers') right to basic well-being, and therefore the right to protect itself from the distracted driver in Figure 1. Bauhn notes that the specifics of these rights have yet to be defined in a real way *(23)*. For example, it is not clear what makes a cost "comparable," and therefore it is unclear how as to the amount of risk a passenger in an automated vehicle can be exposed in order to protect another. This threshold is likely somewhat less than the risk to the pedestrian, but the precise level has not been defined. Vehicle automation allows a fairly precise definition of the levels of acceptable risk, and defining these levels may prove exceptionally difficult.

Descriptive Ethics

An alternative to normative ethics, descriptive ethics is the study of individuals' or groups' beliefs about morality. Unlike normative ethics, any system of ethics based on society's expressed beliefs must be expressed as a distribution. Vehicles may be programmed to respond to the distracted driver collision in Figure 1 probabilistically, to represent the range of society's preferences. While this approach may be easier to defend as it is merely reflecting society's expressed beliefs, it may allow behaviors that, while socially accepted, could still be considered morally wrong. As an example from transportation, survey respondents have assigned lower values-of-life for older pedestrians *(24)*. If an automated vehicle were to factor these values into its collision decisions, it would violate the Institute of Electrical and Electronics Engineers Code of Ethics prohibiting age discrimination *(25)*.

There need not be a single ethical theory that satisfies all. Hansson has noted that radiation exposure guidelines employ the three main ethical theories in the forms of justification (virtue), optimization (utilitarianism/consequentialism), and individual dose limits (deontology) *(26)*. Sandberg and Bradshaw-Martin have argued that automated vehicles are moral proxies for their owners and should therefore reflect their owners' individual morals as closely as possible *(12)*. They propose that owners should be free to select and change their vehicles' moral principles, although the authors acknowledge that this requires an understanding of moral philosophy beyond that of the typical car-buyer. Given that consumers are often guided by inertia in similar situations—such as selecting options for a retirement savings plan *(27)*—whichever ethical theory comes "standard" in new models may unintentionally become the dominant theory.

Potential for Abuse

There are two ways to ensure that an automated vehicle behaves in a way that is morally acceptable. In morally ambiguous situations, an automated vehicle can compare its current situation with a database of similar situations, and perform the recommended action. Given that an automated vehicle can never have perfect situational awareness or predictive abilities, individual scenarios would be needed corresponding to various levels of certainty. This strategy requires some person or group to specify the moral action for individual scenarios. Explicitly defining a vehicle's response encourages transparency in decision making, but may allow a vehicle to respond inappropriately in an unanticipated event where a response has yet to be defined.

An alternative approach is to use machine learning techniques, in which a vehicle is trained with a set of scenarios and responses. The morality of each potential response is scored by a human, and the computer learns from this data to classify different responses as moral or immoral. Machine learning techniques have been success in optical character recognition, language translation, and vehicle automation *(28)*. In contrast to the first approach, machine learning techniques allow a vehicle to respond to novel situations, but the decision-process used to select its response may be difficult to trace back.

Regardless of whether ethical responses are defined explicitly or using machine learning, the exact mechanics of any decision may never be known to the public or government. A lookup table of ethical responses relies on calibrated sensors and vehicle dynamics unique to each model. These tables may be treated either as proprietary information or, like much modern software, as too complex to be understood by any single person. Similarly, many machine learning techniques are "black boxes" where the resulting decision cannot be traced back to its underlying mechanisms. It may be difficult to even recreate its decision, as its mechanisms are often largely probabilistic.

Understandably, the owner of an automated vehicle may have a strong incentive to maximize the safety of his own vehicle and its occupants over the safety of other roadway users. Some may go further, and reject any unnecessary risk to their vehicle. Even if laws are introduced specifying under what conditions an automated vehicle has a duty to rescue, an automaker could design the moral component of the software to be more cautious and risk-averse. Automakers might be unable to advertise or demonstrate their ability to self-protect, but vehicle models may develop reputations for being more risk-averse than others. Due to the complexity of the underlying software, and the difficulty to trace back the decision process behind a vehicle's actions, it may be difficult to ever prove that a vehicle is intentionally avoiding risk. Testing may be needed to ensure compliance, performed either by government or through industry self-regulation.

CONCLUSIONS

The automation of road vehicle introduces several new problems, including the need for some type of moral reasoning, either by engineers when developing crash avoidance strategies, or encoded directly in the vehicle's own path planning algorithms. A particularly difficult moral problem is determining when an automated vehicle must subject itself (and its passengers) to a small risk in order to greatly reduce the risk of others. We have shown that an automated vehicle programmed to foremost protect the safety of its own passengers can produce morally unacceptable results. Common law does not require intervention in an emergency in most cases, even if there is no risk for the potential rescuer. The ethics literature provides a language for discussing these problems, and several possible solutions. Should society decide that advanced automated vehicles should occasionally subject their occupants to small levels of avoidable risk in order to protect other users, regulation is needed to ensure that industry does not hide excessive self-protection tendencies within complex software. This article has defined this problem and discussed initial directions.

REFERENCES

1. Levinson, J., J. Askeland, J. Becker, J. Dolson, D. Held, S. Kammel, J. Z. Kolter, D. Langer, O. Pink, V. Pratt, M. Sokolsky, G. Stanek, D. Stavens, A. Teichman, M. Werling, and S. Thrun. Towards Fully Autonomous Driving: Systems and Algorithms. Presented at the 2011 IEEE Intelligent Vehicles Symposium (IV), 2011.
2. Ferguson, D. I. F., and D. A. Dolgov. Modifying Behavior of Autonomous Vehicle Based on Predicted Behavior of Other Vehicles, United States Patent Application 20130261872 Kind Code: A1,Oct 03, 2013.
3. NHTSA. *National Motor Vehicle Crash Causation Survey*. Publication DOT HS 811 059. National Highway Traffic Safety Administration, Washington, D.C., Jul. 2008.

4. Anderson, J. M., K. Nidha, K. D. Stanley, P. Sorenson, C. Samaras, and O. A. Oluwatola. *Autonomous Vehicle Technology: A Guide for Policymakers*. Publication RR-443-RC. RAND Corporation, Santa Monica, CA, Jan. 2014.
5. National Highway Traffic Safety Administration. *Preliminary Statement of Policy Concerning Automated Vehicles*. Publication NHTSA 14-13. National Highway Traffic Safety Administration, Washington DC, May 2013.
6. Llaneras, R. E., J. A. Salinger, and C. A. Green. Human Factors Issues Associated with Limited Ability Autonomous Driving Systems: Drivers' Allocation of Visual Attention to the Forward Roadway. Presented at the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, Bolton Landing, NY, 2013.
7. Gold, C., D. Damböck, L. Lorenz, and K. Bengler. "Take Over!" How Long Does it Take to Get the Driver Back Into the Loop? In *Human Factors and Ergonomics Society Annual Meeting*, No. 57, San Diego, CA, 2013, pp. 1938–1942.
8. Paul, R. 7 Cool or Disturbing Things About High-Tech Cars. Nov. 20, 2013. http://www.consumerreports.org/cro/news/2013/11/7-cool-or-disturbing-things-about-high-tech-cars/index.htm. Accessed Dec. 2, 2013.
9. Gabler, H. C., and W. T. Hollowell. The Crash Compatibility of Cars and Light Trucks. *Journal of Crash Prevention and Injury Control*, Vol. 2, No. 1, Mar. 2000, pp. 19–31.
10. Goodall, N. J. Ethical Decision Making During Automated Vehicle Crashes. *Transportation Research Record: Journal of the Transportation Research Board*, forthcoming 2014.
11. Goodall, N. J. Machine Ethics and Automated Vehicles. In *Road Vehicle Automation* (G. Meyer and S. Beiker, eds.), Springer International Publishing, pp. 93–102.
12. Sandberg, A., and H. Bradshaw-Martin. Autonomous Vehicles, Moral Agency and Moral Proxyhood. Presented at the Beyond AI: Artificial Golem Intelligence, Pilsen, Czech Repulblic, 2013.
13. Bilger, B. Auto Correct: Has the Self-Driving Car at Last Arrived? The New Yorker, Nov 25, 2013.
14. Smith, B. W. *Automated Vehicles are Probably Legal in the United States*. Center for Internet and Society at Stanford Law School and Center for Automotive Research at Stanford, Nov. 2012.
15. Chopra, S., and L. F. White. *A Legal Theory for Autonomous Artificial Agents*. University of Michigan Press, Ann Arbor, 2011.
16. Groninger, J. No Duty to Rescue: Can Americans Really Leave a Victim Lying in the Street? What Is Left of the American Rule, and Will It Survive Unabated? *Pepperdine Law Review*, Vol. 26, No. 2, Oct. 2012.
17. Grush, E. H. The Inefficiency of the No-Duty-to-Rescue Rule and a Proposed "Similar Risk" Alternative. *University of Pennsylvania Law Review*, Vol. 146, No. 3, Mar. 1998, pp. 881–900.
18. Levit, N. The Kindness of Strangers: Interdisciplinary Foundations of a Duty to Act. *Washburn Law Journal*, Vol. 40, No. 3, 2001.
19. Scordato, M. R. Understanding the Absence of a Duty to Reasonably Rescue in American Tort Law. *Tulane Law Review*, Vol. 82, No. 4, 2008, pp. 1447–1503.
20. Carter, M. C. Morality, Law and the Duty to Act: Creating a Common Law Duty to Act Modeled After the Responsibility to Protect Doctrine. *Washington University Jurisprudence Review*, Vol. 2, No. 1, Jan. 2010, pp. 138–164.
21. Sinnott-Armstrong, W. Consequentialism. In *The Stanford Encyclopedia of Philosophy* (E. N. Zalta, ed.), Winter 2012.

22. Alexander, L., and M. Moore. Deontological Ethics. In *The Stanford Encyclopedia of Philosophy* (E. N. Zalta, ed.), Winter 2012.
23. Bauhn, P. The Extension and Limits of the Duty to Rescue. *Public Reason*, Vol. 3, No. 1, Jun. 2011, pp. 39–49.
24. Johansson-Stenman, O., and P. Martinsson. Are Some Lives More Valuable? An Ethical Preferences Approach. *Journal of Health Economics*, Vol. 27, No. 3, May 2008, pp. 739–752.
25. Institute of Electrical and Electronics Engineers. The IEEE Code of Ethics. 1963. http://www.ieee.org/about/corporate/governance/p7-8.html. Accessed Apr. 25, 2014.
26. Hansson, S. O. *The Ethics of Risk: Ethical Analysis in an Uncertain World*. Palgrave Macmillan, 2013.
27. Madrian, B. C., and D. F. Shea. The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior. *The Quarterly Journal of Economics*, Vol. 116, No. 4, Nov. 2001, pp. 1149–1187.
28. Batavia, P., D. Pomerleau, and C. Thorpe. *Applying Advanced Learning Algorithms to ALVINN*. Publication CMU-RI-TR-96-31. Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, Oct. 1996.