Ethical Decision Making in Autonomous Vehicles During Unavoidable Crashes

Engineering
School of Engineering and Applied Sciences
Department of Civil and Environmental Engineering
Noah Joseph Goodall, Ph.D. Candidate, Smith, BL and Park, B.

Autonomous Vehicles Will Still Crash

Recent advances in passenger vehicle automation have received much attention in the media, most notably Google's self-driving car. There has been little discussion of the behavior of these vehicles when a crash is unavoidable. Some assume that a well-functioning autonomous vehicle will never crash, an unrealistic expectation given limited maneuverability at freeway speeds. Others assume that a human driver will continuously monitor the roadway and to avoid crashes, but early research shows that even first-time drivers in autonomous vehicles are often inattentive.

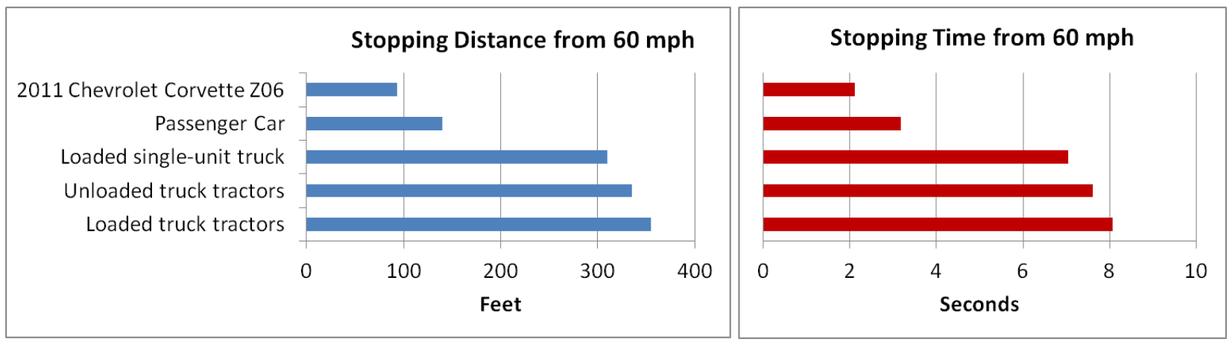
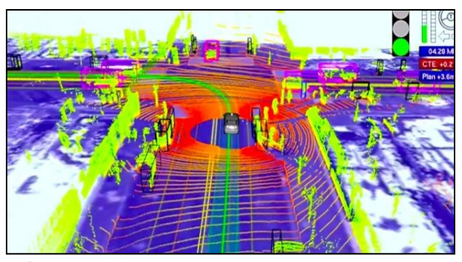



photo by Jurvetson (flickr)


Source: IEEE


Source: Jalopnik

Crashing is Complicated

When autonomous vehicles crash, unlike other autonomous transportation such as personal rapid transit or an aircraft's autopilot, they must decide how best to crash. This is an exceptionally complex task which requires the vehicle to make subtle moral decisions. For example, the autonomous vehicle in Figure 1 has three path options after a bus drifts into its lane, each with complicated repercussions.

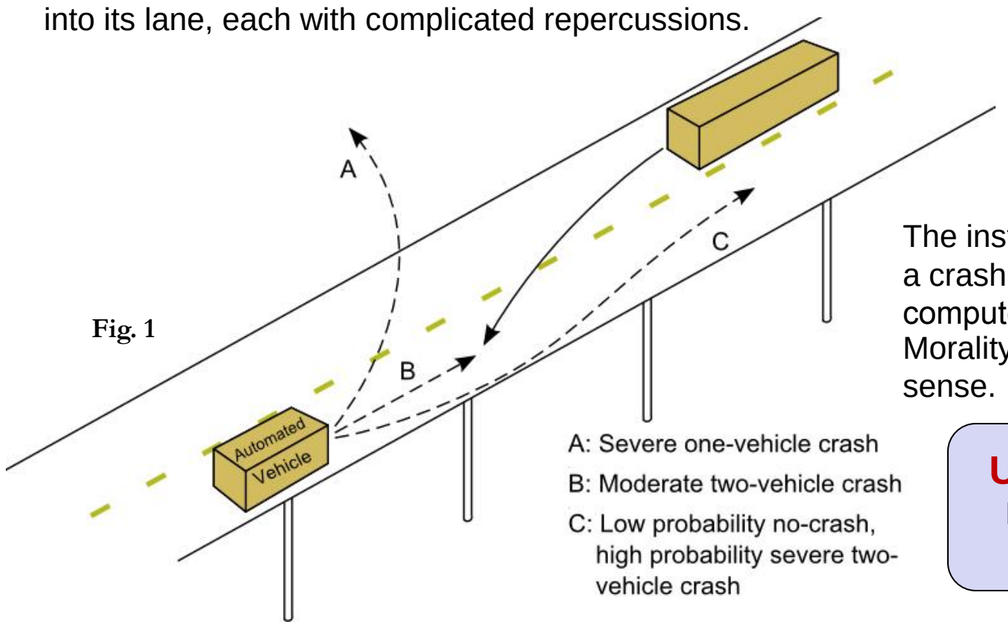


Fig. 1

A: Severe one-vehicle crash
B: Moderate two-vehicle crash
C: Low probability no-crash, high probability severe two-vehicle crash

Shortcomings of Rule-based Systems

The instinct for engineers is to code a set of behavior rules. In a crash, any rule-based moral system will struggle with computers' literalness. Morality requires common sense.

Utilitarianism
Minimize global damage

Literal Interpretation
- Given a choice, crashes into vehicle with higher safety rating.
- Uses insurance industry damage estimates and avoids collisions with expensive vehicles.
- May protect other cars first, putting its own passengers at greater risk.

Asimov's Three Laws of Robotics
1. Do not injure human or let them come to harm through inaction
2. Follow human's order, unless it conflicts with First Law
3. Do not harm self, unless this conflicts with First or Second Law

Literal Interpretation
- Refuses to drive above 20 miles per hour.
- Won't brake heavily to avoid a collision (causes whiplash).

Rules will conflict | Rules are unclear | Unintended results

This research investigates issues in ethical decision making in autonomous vehicles from findings in philosophy, artificial intelligence, and robotics. The following three-phase approach is proposed, to be enforced as technology become available.

Phase 1: Rule-based
1. "Top-down" approach
2. Develop safety metric, independent of insurance costs
3. Vehicle tries to maximize utility
4. If unsure, decelerate and evade

Phase 2: Common Sense
1. "Bottom-up" approach
2. Machine learning of driving ethics
3. Trained by a combination of simulation of near-crashes, general rules, and human feedback.

Phase 3: Feedback

Autonomous vehicle defends its actions using natural language

Quality

Autonomous vehicles, most notably the Google self-driving car, are quickly becoming a reality. These vehicles will still occasionally crash, often before the driver can be alerted and take over. A computer must decide how the vehicle should crash. In situations where injury is unavoidable this becomes a complex ethical decision.

Impact

34,000 Americans died in vehicle crashes last year. Autonomous vehicles may greatly reduce this number, but only if drivers are comfortable with the technology. Ethical vehicles are essential for driver acceptance.

Advance in Knowledge

The crashing behavior of autonomous vehicles has never been studied. This work assists vehicle developers and provides guidance to legislators unfamiliar with how to regulate these vehicles.

Innovation

This will be the first project to study machine ethics in a civilian transportation application.

Interdisciplinary

Combines advanced issues in philosophy, robotics, artificial intelligence, and transportation.

Virginia Center for Transportation INNOVATION & RESEARCH
UNIVERSITY of VIRGINIA

Annual Presidential Research Poster Competition
May 3, 2013