# Ethical Decision Making in Automated Vehicles During Unavoidable Crashes

Noah J. Goodall, Ph.D., P.E.
Virginia Center for Transportation Innovation and Research

Virginia Center *for* Transportation INNOVATION & RESEARCH

VDOT Virginia Department of Transportation

Transportation Research Board 2nd Annual Workshop on Road Vehicle Automation – July 16, 2013

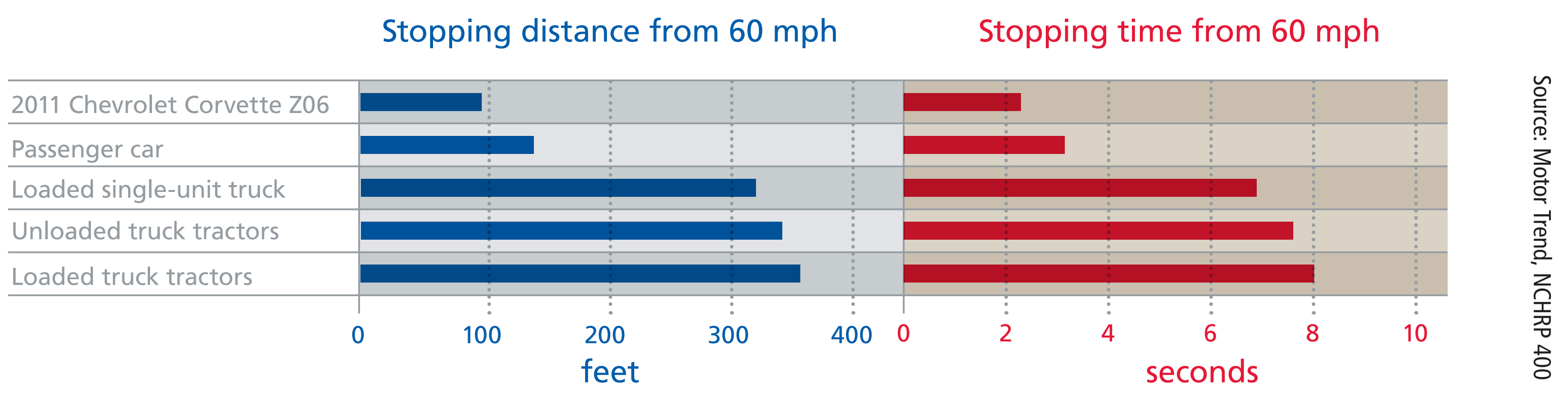
## Automated Vehicles Still Will Crash

Automated vehicles are quickly becoming a reality. On-road testing is legal in several states.

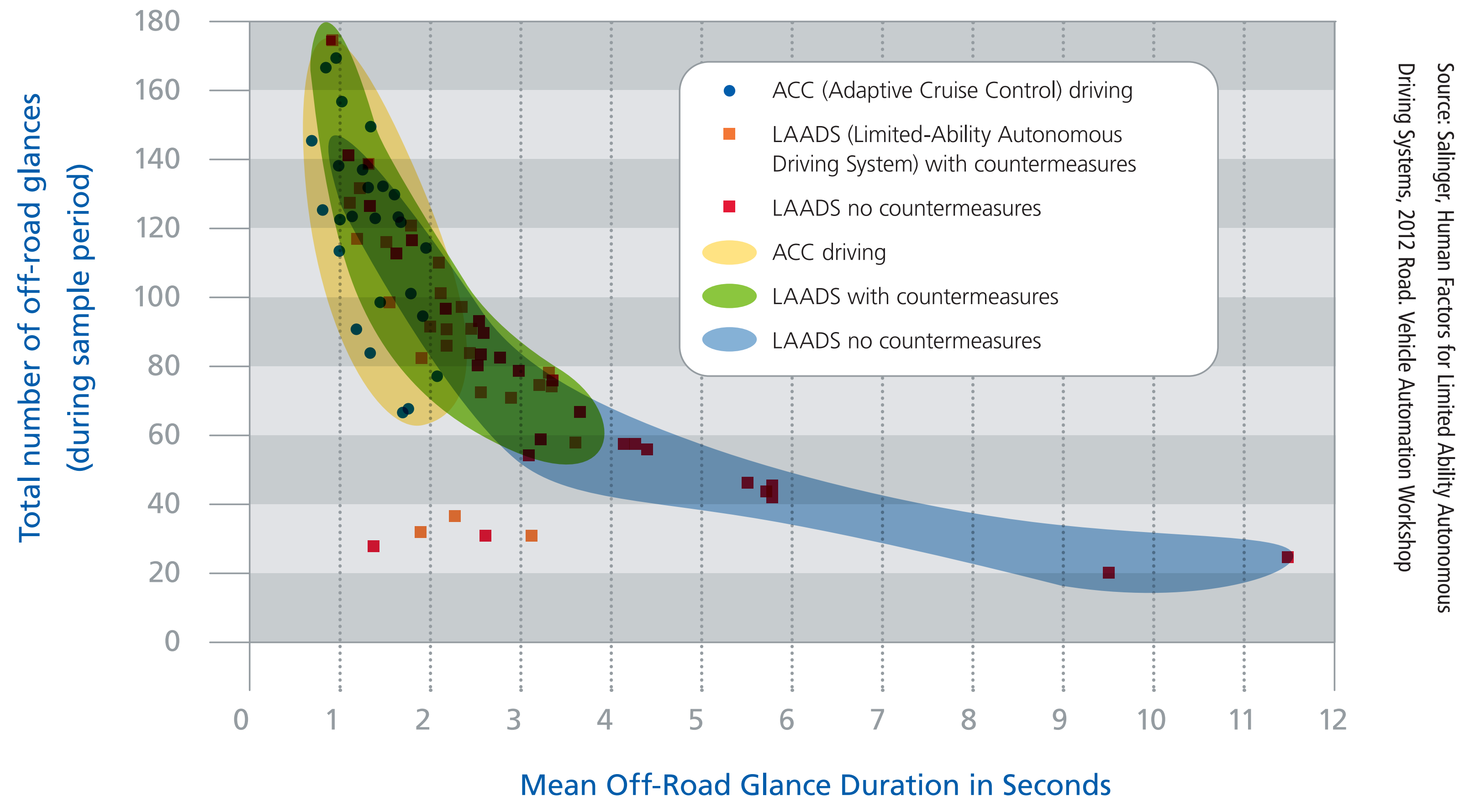
However, there has been little discussion of the behavior of these vehicles when a crash is unavoidable.

Some assume a well-functioning automated vehicle will never crash. This is an unrealistic expectation, given the limited maneuverability at freeway speeds of such a vehicle and the unpredictability of other vehicles, pedestrians, cyclists, wildlife and debris.

Others assume a human driver will continuously monitor the roadway to avoid crashes, but early research shows even first-time drivers in automated vehicles often are inattentive.

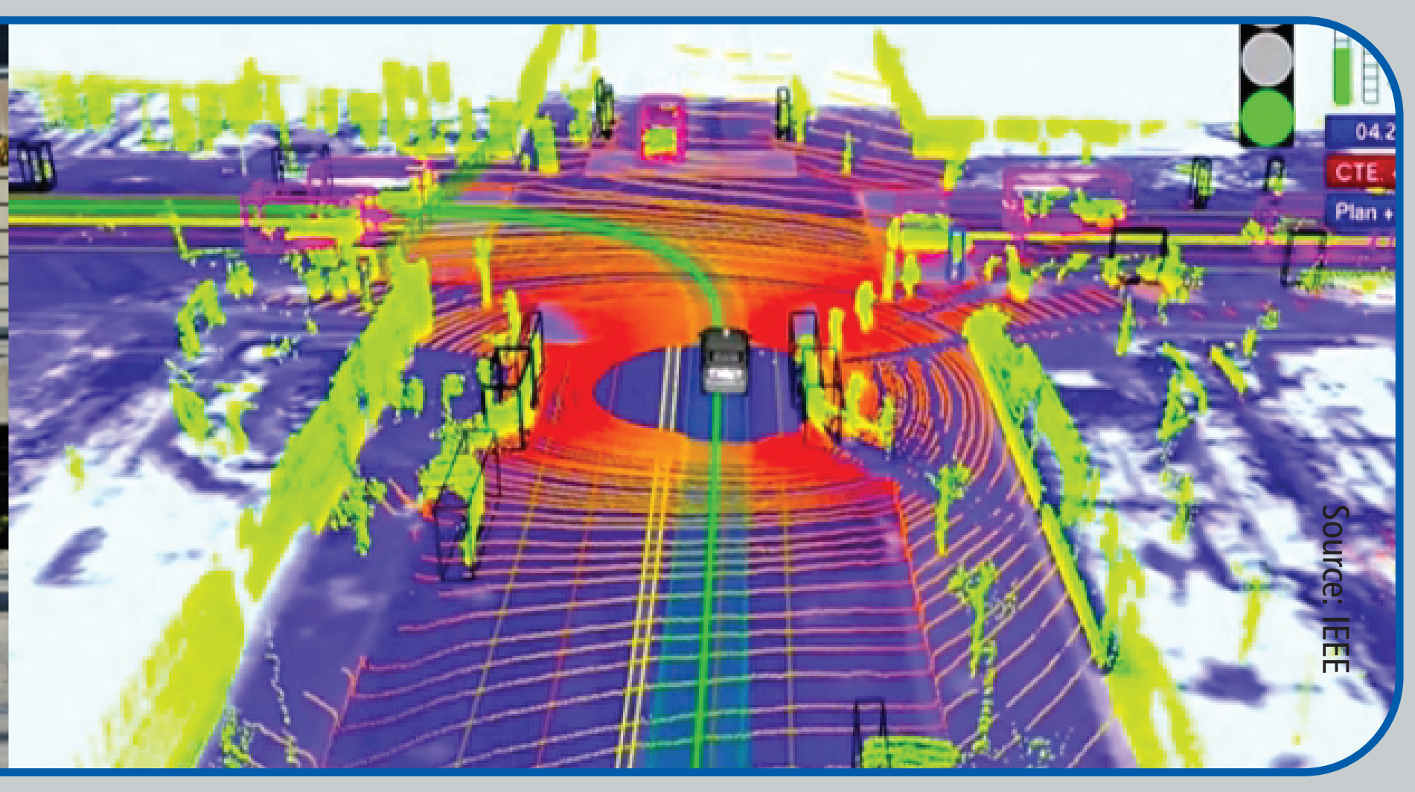


Source: Motor Trend, NCHRP 400



Source: Salinger, Human Factors for Limited Ability Autonomous Driving Systems, 2012 Road. Vehicle Automation Workshop

photo by Jurvetson (flickr)

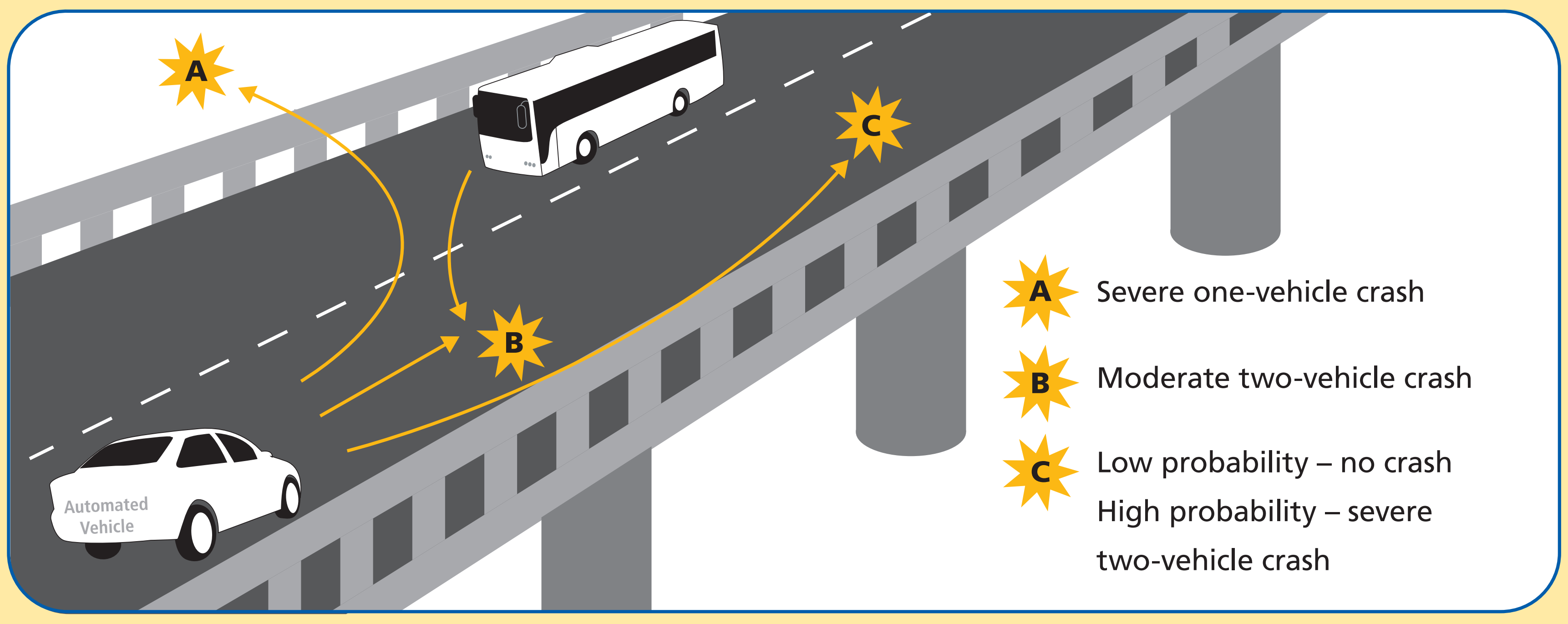


Source: IEEE

## Crashing is Complicated

Unlike other automated vehicles – such as aircraft, where every collision is catastrophic, and guided track systems, which can only avoid collisions in one dimension – automated vehicles on a roadway can evaluate different pre-crash trajectory alternatives and select a path with the lowest damage or likelihood of collision.

This is an exceptionally complex task that requires the vehicle to make subtle moral decisions. For example, the automated vehicle has three path options after a bus drifts into its lane, each with complicated repercussions.



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 the computer's literalness. Morality requires common sense.

### Asimov's Three Laws of Robotics

1. Do not injure humans 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)

### 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

**Rules will conflict • Rules are unclear • Unintended results**

## Proposed Approach

This research investigates issues in ethical decision making in automated vehicles from findings in philosophy, artificial intelligence and robotics.

The following three-phase approach is proposed, to be enforced as technology becomes 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 and recordings of near-crashes, the rule-based system from Phase 1, and human feedback.

### Phase 3: Feedback

1. Automated vehicle defends its actions using natural language
2. Mistakes can be understood and corrected