

The Dilemma of Teaching Ethics to Self-Driving Cars

By Doug Newcomb

ARTICLE DATE : October 9, 2015

pcmag.com

It didn't take long [riding in Google's self-driving car last week](#) to realize that the technology is not only ready for the road, but in some instances makes better decisions than human drivers—slowing and yielding to bicyclists, for example.



Opinions

But in order to do more than just creep along city streets, autonomous technology will have to make the kind of critical decisions that are second nature to experienced human drivers. And it may have to negotiate ethical dilemmas as much as tricky driving situations.

As an example, self-driving technology adds a new twist to the century-old philosophical dilemma known as the "[trolley problem](#)." In this scenario, a person has to decide whether to pull a lever at a Y intersection and drive over one person who is tied to the track in order to save five people tied to the adjacent track.

For self-driving cars, this has been recast as the "tunnel problem." Imagine that an autonomous vehicle is traveling on a single-lane mountain road and about to enter a tunnel, when a child inadvertently crosses into its path just inside the entrance so that the car has to make a split-second decision. Does it continue straight and hit the child? Or does it swerve and hit the tunnel, injuring or killing the car's occupants?

The day after getting a ride in Google's self-driving car in Mountain View, California, I attended an event at Mercedes-Benz's North American R&D facility in nearby Sunnyvale. Among several topics covered throughout the day, Stanford professor and head of the university's [Revs program](#) Chris Gerdes gave a presentation that delved into the subject of ethics and autonomous cars.

Gerdes revealed that Revs has been collaborating with Stanford's philosophy department on ethical issues involving autonomous vehicles, while the university has also started running a series of tests to determine what kind of decisions a robotic car may make in critical situations.

When Philosophy Takes the Wheel

As part of his presentation, Gerdes made a case for why we need philosophers to help study these issues. He pointed out that ethical issues with self-driving cars are a moving target and "have no limits," although it's up to engineers to "bound the problem."

To do this and move the ethics of self-driving technology beyond a mere academic discussion, Revs is running experiments with Stanford's [x1 test vehicle](#) by placing obstacles in the road. He noted that placing different priorities within the vehicles' software program have led to "very different behaviors."

Gerdes asserted that this nascent ethics-based programming of self-driving software could become a "core requirement" for the technology—and not something just discussed in ivory towers. And he added that ethics isn't something that automakers can source from a "Tier 1" supplier like, say, a fuel pump or in-dash screen. The recent [VW Dieselgate scandal](#) has certainly shown that.

Experienced human drivers have been programmed for years to deal with split-second decisions—and they still don't always make the right ones. But after seeing how Google's self-driving vehicles react to everyday decisions, and hearing about the work that Stanford Revs and the school's philosophy department is conducting, I'm betting that the cars will ultimately make smarter decisions.

