The Washington Post

Business

Will the public accept the fatal mistakes of self-driving cars?

By Steven Overly February 20

How many people could self-driving cars kill before we would no longer tolerate them?

This once-hypothetical question is taking on greater urgency, particularly among policymakers here in Washington. The promise of autonomous vehicles is that they will make our roads safer and more efficient, but no technology is without its shortcomings and unintended consequences — in this instance, potentially fatal consequences.

“What if we can build a car that’s 10 times as safe, which means 3,500 people die on the roads each year. Would we accept that?” asks John Hanson, a spokesman for the Toyota Research Institute, which is developing the automaker’s self-driving technology.

“A lot of people say, ‘If I could save one life, it would be worth it.’ But in a practical manner, though, we don’t think that would be acceptable,” Hanson adds.

Members of Congress are beginning to consider legislation meant to enable the broader adoption of self-driving technology without compromising safety. At a House subcommittee hearing last week, for example, lawmakers and industry leaders alike grappled with the question of whether machines need only

drive better than humans to win our trust.

The Transportation Department, for its part, published its first guidelines for self-driving vehicles last year in an effort to keep pace with automakers that hope to unleash the cars on the road in the next several years. Ford, for example, has set a goal of releasing an autonomous vehicle fleet by 2021.

More than 35,000 people were killed in car crashes in the United States in 2015, according to the National Highway Traffic Safety Administration. The agency estimates 94 percent of those crashes were the result of human error and poor decision-making, including speeding and impaired driving.

Self-driving enthusiasts say that the technology could make those deaths a misfortune of the past. But humans are not entirely rational when it comes to fear-based decision-making. It's the reason people are afraid of shark attacks or plane crashes, when the odds of either event are exceptionally low.

Calestous Juma, a Harvard University professor and an expert in technology and sustainable development, draws a parallel between self-driving cars and home refrigerators, which gained popularity in U.S. households in the 1920s and 30s. Although scientists understood that cold storage could cut down on food-borne illnesses, reports of refrigeration equipment catching fire or leaking toxic gas made the public wary.

Americans eventually adopted the now-ubiquitous household appliance, thanks in large part to the Agriculture Department, which advocated for the health benefits of refrigeration and explained the technology's safety, Juma writes in his book, "Innovation and Its Enemies: Why People Resist New Technologies."

People are also more inclined to forgive mistakes made by humans than machines, Gill Pratt, chief executive of the Toyota Research Institute, told lawmakers on Capitol Hill last week.

"The artificial intelligence systems on which autonomous vehicle technology will depend are presently and unavoidably imperfect," Pratt told lawmakers at a House subcommittee hearing last week. "So, the question is 'how safe is safe enough' for this technology to be deployed."

As a society, we understand human limitations because we live with them daily, said Iyad Rahwan, an associate professor at the Massachusetts Institute of Technology Media Lab who has studied social dilemmas presented by autonomous vehicles. While we may assign blame or seek retribution — by sending a drunk driver to prison, for example — the capacity for human failure is not hard to understand or empathize with. The same is not true for machines, he said.

“We penalize them and distrust them more when they make mistakes,” Rahwan said. “It comes down to us not having proper mental models of what machines can and cannot do.”

Researchers at the University of Pennsylvania have dubbed this “algorithm aversion.” In a 2014 study, participants were asked to observe a computer and human make predictions about the future, such as how a student would perform based on past test scores. Researchers found that “people more quickly lose confidence in algorithmic than human forecasters after seeing them make the same mistake.”

The answer to questions about safety may therefore come down to how much we trust self-driving cars, regardless of how many lives they can save, Rahwan said. For example, if autonomous vehicles save the lives of thousands of motorists but cause fatalities of cyclists and pedestrians to increase, the public’s trust in the technology is likely to erode.

“If they’re not comfortable with the trade-offs that cars are making, then we risk people losing faith in the system and perhaps not adopting the technology,” Rahwan said.