Self-driving cars don't care about your moral dilemmas

As self-driving cars move from fiction to reality, a philosophical problem has become the focus of fierce debate among technologists across the world. But to the people actually making self driving cars, it's kind of boring.

The "trolley problem" is the name for a philosophical thought experiment created as an introduction to the moral distinction between action and inaction. The classic example is a runaway mine cart, hurtling down tracks towards a group of five oblivious people. With no time to warn them, your only option is to pull a switch and divert the cart on to a different track, which only has one person standing on it. You will save five lives, but at the cost of actively killing one person. What do you do?

All kinds of tweaks and changes can be made to the basic problem, to examine different aspects of moral feeling. What if, rather than pulling a switch, you stop the mine cart by pushing one particularly large person in its way? What if the five people are all over the age of 80 and the one person is under 20? What if the five people are in fact five hundred kittens?

What if rather than a mine cart, you were in a runaway self-driving car? What if, rather than making the decision in the heat of the moment, you were a programmer who had to put your choices into code? And what if, rather than picking between the lives of five people and one person on different roads, you had to pick between the life of the car's sole occupant, and the lives of five pedestrians?

It seems like a question that cuts to the heart of fears over self-driving cars: putting questions of life and death in the hands of coders in California who make opaque decisions that may not be socially optimal. After all, would you buy a car if you knew it was programmed to swerve into a tree to protect someone who crossed the road without looking?

A recent paper in the journal Science suggested that even regulation may not help: polling showed that regulation mandating such self-sacrifice wouldn't be supported by

a majority of people, and that they'd avoid buying self-driving cars as a result. That, of course, would result in far more deaths in the long run, as the endless deaths at the hands of incapable human drivers would continue.

But to engineers at X, the Google sibling which is leading the charge to develop fully self-driving cars, the questions aren't as interesting as they sound. "We *love* talking about the trolley problem", joked Andrew Chatham, a principal engineer on the project.

"The main thing to keep in mind is that we have yet to encounter one of these problems," he said. "In all of our journeys, we have never been in a situation where you have to pick between the baby stroller or the grandmother. Even if we did see a scenario like that, usually that would mean you made a mistake a couple of seconds earlier. And so as a moral software engineer coming into work in the office, if I want to save lives, my goal is to prevent us from getting in that situation, because that implies that we screwed up.

"It takes some of the intellectual intrigue out of the problem, but the answer is almost always 'slam on the brakes'," he added. "You're much more confident about things directly in front of you, just because of how the system works, but also your control is much more precise by slamming on the brakes than trying to swerve into anything. So it would need to be a pretty extreme situation before that becomes anything other than the correct answer."

Even if a self-driving car did come up against a never-before-seen situation where it did have to pick between two accidental death scenarios, and even if the brakes failed, and even if it could think fast enough for the moral option to be a factor (Nathaniel Fairfield, another engineer on the project, jokes that the real question is "what would you ...oh, it's too late"), there remains no real agreement over what it should do even in idealised circumstances. A public tool released alongside the Science paper allows individuals to create their own ethical dilemmas, and the only consistent finding is that people are inconsistent – even when it comes to their own views. So it's probably for the best that we aren't trying to code those views into our cars just yet.