

6. Trolley Problems?

The trolley problem was introduced in 1967 by philosopher Phillipa Foot at Oxford University to defend the doctrine of double-effect by testing which kinds of intentions pre-theoretically seem to matter to us. Its two best-known versions were formulated by philosopher Judith Jarvis Thomson at Massachusetts Institute of Technology.³⁴ It has since been re-imagined by so many philosophers and instructors that there are now almost countless variations. The classic formulation presents you with a moral dilemma: a runaway street-car is about to run over five oblivious railroad workers, but by pulling a lever you can change its course so that it only runs over one worker. It is, in other words, a type of Kobayashi Maru; there are no good choices, and the point of the experiment is not to test your ability to choose correctly, but to provide insight into our strongest intuitions about what is ultimately morally required.

Critics of using trolley problems to teach ethics or morality argue that the situations are horrific, unrealistic, and teach students little about morality or moral decision-making in the real world. For example, Brianna Rennix and Nathan J. Robinson write, “It’s not just that, as the additional conditions grow, there are not any obvious right answers. It’s that every single answer is horrific, and wild examples like this take us so far afield from ordinary moral choices that they’re close to nonsensical.”³⁵ No matter what you do, even if you do nothing at all, death is certain, and you are either directly or indirectly responsible for that death. Since few (if any) students believe they will ever find themselves in situations even closely resembling such fantastical and macabre dilemmas, teachers often find that the exercise is met with ridicule rather than serious consideration. Critics of trolley problems argue that students would gain a better understanding of morality using examples that are applicable to their lives.

On the other hand, supporters of trolley problems claim that they are a useful tool for teaching students how to examine their own moral intuitions and distinguish the differences between two popular moral theories: utilitarianism and deontology. If students decide to sacrifice one railroad worker to save five, teachers often say they are focusing on the consequences and the minimization of suffering, thereby making a utilitarian calculation. If they instead argue that killing is always wrong, even if it is to save more lives, then teachers often suggest they are adhering to deontological justifications where adherence to duty is more important than consequences.

One troubling extension of the way that trolley problems are often used in the classroom is the suggestion that similar thought experiments may be used to set rules in real life. Researchers at MIT have created an online interactive trolley problem style website that gathers data about how people report they would decide a range of difficult moral dilemmas. Some have suggested that this data could be used to implement human-style morality in autonomous machines such as self-driving cars. This proposed application is disconcerting in part because the people taking the

³⁴ Lauren Cassani Davis, “Would You Pull the Trolley Switch? Does it Matter?” The Atlantic, October 9, 2015, <https://www.theatlantic.com/technology/archive/2015/10/trolley-problem-history-psychology-morality-driverless-cars/409732/>

³⁵Brianna Rennix and Nathan J. Robinson, “The Trolley Problem will tell you nothing useful about morality,” Current Affairs, November 3, 2017 <https://www.currentaffairs.org/2017/11/the-trolley-problem-will-tell-you-nothing-useful-about-morality>

online quiz in a spare moment surely are not grappling with the significance of their choices in a robust way. Further, this sort of self-report, especially of hypothetical values, is highly unreliable. It is also far from obvious that a democratic approach to the morality of autonomous vehicles is desirable—perhaps some principles should override the wishes of the majority.

Referring to the newest application of the old thought experiment, Lauren Cassani Davis from *The Atlantic* writes, “[a]s human agents are replaced by robotic ones, many of our decisions will cease to be in-the-moment, knee-jerk reactions. Instead, we will have the ability to premeditate different options as we program how our machines will act...this is the perfect example of where theory collides with the real world—and thought experiments like the trolley problem, though they may be abstract or outdated, can help us to rigorously think through scenarios before they happen.”³⁶ Unfortunately, it is far from clear that trolley problem thought experiments help us to think rigorously about real world situations.

³⁶ Davis, *supra* note 34