

Should autonomous vehicles be programmed to respond in trolley-like situations using a cultural relativist framework? Why or why not?

Autonomous vehicles (AV) are among us. With components smaller than a penny controlling the entire vehicle's climate control system, there is a lot to make of how AVs will evolve the perception of a vehicle. Starting out as living animals that were limited in speed and energy; designers developed mechanical machines. As technologies continue to evolve, many of these mechanical machines make decisions for their drivers, using Artificial Intelligence, which is derived from the way the human brain makes decisions. The AVs decide how centered the user is in the lane, which routes to take to a destination, and when to decelerate. These decision-making features reduce human intervention, but at what cost? It is now up to the AV to decide what to do when there is an imminent collision. It is important to consider that prior to the existence of AVs, it was up to the driver what to do in the case of an imminent collision, the brain made the split-second decision. Therefore, fitting the vehicle's decision-making processor with a cultural relativist framework would be most consistent with the status quo of driving conventions today.

The cultural relativist framework expresses that each culture and belief is valuable and unique such that it only should be understood and analyzed from its own perspective. This is very similar to how any given individual's personality, beliefs, and morals work. As individuals, we have our own perspectives on morals that do not extend one to one with anyone else in the world. There may be similarities among people, but no two people share the exact same set of beliefs for every dilemma.

Similar to a fingerprint, there are subtle differences among every person which is what makes us all unique. In a non-autonomous vehicle, the driver controls the decisions of the vehicle. In other words, the cars are fitted with a human that makes decisions based on a cultural relativist framework. In the event of an oncoming collision, it is up to the driver to decide how to react. Depending on who is driving, they may choose to sacrifice themselves or save themselves. There is no way to generalize what the decision would be for all drivers on Earth and come to a definite answer for everyone. There is too much variety.

In order to apply a cultural relativist framework to the vehicle, the vehicle would need to use its processing power to decide, “What would the driver do in the situation at hand?”. When that is known – or at least predicted with a certain threshold of confidence – then the vehicle would be theoretically identical to a non-autonomous vehicle. In other words, if the AV learns enough about the driver to the point to which it knows what the driver will decide, then it would be indistinguishable with a non-autonomous vehicle. From this point of view, it is clear that the cultural relativist framework is a convincing tool for automating the decisions of the vehicle without over-analyzing a problem that has no solution. Therefore, when developing a decision-making processor for an autonomous vehicle, using a cultural relativism framework would likely result in the least amount of change in terms of how current driving habits apply in a real-world setting.

Universalism, the opposite of relativism, could be argued to be the most fair framework to implement in the AV. The backing of universalism is that there is a truth that applies to everyone and that all rewards and consequences are equally distributed to everyone in the world. The trolley problem and the universalist system do not blend well, since one conceivable truth is “no lives should be lost at the expense of another”, however, for the sake of argument this framework would likely prioritize saving more humans over less. If the vehicle were to apply this framework, it would most likely follow a static list-checking procedure whenever a collision is imminent – rather than a dynamic decision-making procedure based on the driver’s preferences – following a list of set “universal truths”, which would be the same in every vehicle. This would be consistent with many of the rules that drivers must follow on the road, such as staying in one lane, cruising at or under the speed limit, etc. Consequently, applying a universalist framework would make the vehicles far more predictable as they would follow a finite set of truths for all scenarios.

Universalism fails to address many political and economic processes. On one hand, what are the universal truths? Does the AI system decide the truths? Do the vehicle manufacturers? Since the truths are the same for everyone, it would be a list, as explained above, that would need to be decided upon somehow. In a society of relative perspectives, beliefs, and systems, the impediments that this system would encounter on its way through a human-driven business would keep it in the prototype phase indefinitely. Further, if the AV did make it to the market, the majority of prospective buyers would likely not like to hear that they would be sacrificed if they did not fit into the

universal truth list, especially since they would not be controlling the AV prior to the crash.

The purpose of a vehicle is to transport its driver from point A to point B. The first vehicle humans encountered were their feet. Their feet listened to what their brain told them. When humans began using animals as their vehicles, the animals also listened to the humans (for the most part), taking the humans where they indicated. In the age of mechanical vehicles, nothing has changed, the vehicle accepts user feedback to go forward, backward, left, and right. Now, we must decide how to create autonomous vehicles, and there is an effort to complicate matters with “big brother” frameworks. The user is at the center of the equation, it is the user who purchased the system, it is the user who made the decisions prior to AVs, and it is the user who is responsible for their actions. The act of translating a person's actions into a framework that the vehicle follows definitely adds complexity to legal dilemmas, but under an ethical and moral scope, the situation does not change. A vehicle that does “what the driver would do” is the path of least resistance, the path of least change, and the way to be as consistent as possible with drivers' experience, in regards to not only the trolley problem but all other driving situations as well. The reality is that AVs are made possible by Artificial Intelligence, which was made possible by studying how the human brain works. Therefore, applying a brain-inspired technology to a problem that – under pure human intervention – would be handled only using the driver's brain, is the most rational way to make a system for a user who expects the vehicle to behave in a certain manner.

Unfortunately, I've said all I have to say and have am 75 words below the word count threshold. I could go back and try to fill in a word here and there or "elaborate" on something that doesn't need elaboration. Instead, I am using this text to say thank you for teaching us a variety of ethics and making us more culturally understanding citizens. I hope you enjoy reading all the diverse responses to this prompt.