PHILOS 3235: Test #2

Due on Wednesday, May 10th, 2017

Dittmer 15:00

Illya Starikov

Monday, April 3rd, 2017

Problem I

> Utilitarianism states that actions are right just in case they produce the most overall (available) goodness. Discuss some ways in which according to Kantian ethics, utilitarianism is false.

Utilitarianism and Kantian ethics seem to have have a large overlap. After all, utilitarianism is the doctrine that actions are right if they benefit the majority and Kantian Ethics reflect much of this thinking. However, when analyzing the two frameworks of ethics, there are areas where they differ. The biggest area of difference is in the first Categorical Imperative.

The first Categorical Imperative states that

> Act only according to that maxim by which you can at the same time will that it should become a universal law.

This may seem like a perfect utilitarian definition; until one realizes that "universal" is not the same as "maximum benefit". To show how this juxtaposition is true, let us take two examples.

One of the most widely known examples of a wrong in Kantian Ethics is suicide. If everyone committed suicide (making it universal), the human-centric ethics would fall apart. However, we *know* there to be situation where self-sacrifice is the correct in a utilitarian framework. Consider the transplant problem.

> Suppose you are an outstanding surgeon working in the top medical facility in your area. During a shift, you have five subsequent patients needing transplants within the hour (with their life in jeopardy). In a stroke of luck, a healthy patient walks is in for a checkup that has all the organs necessary. Do you save five at the expense of one?

The one patient is **obligated** to sacrifice himself to save five other patients. This is not the case in a Kantian framework.

Take another example: a situation where you are forced to lie. Imagining back to Nazi Germany in WWII, no sane human being would admit to harboring a fugitive Jew. In a utilitarian framework, not getting the innocent person killed would definitely be considered the maximum benefit. However, in pure Kantian Ethics, lying is very much breaking the first Categorical Imperative.

Admittedly, there is a lot of overlap in Utilitarian and Kantian; most of which is at the intersection of maximum benefit and universal benefit. However, from the point of maximum (whether it be at the 51% mark or on) to the point of universal (which has to lie at the 100% mark), there is a tangible difference.

Problem II

> Utilitarianism states that actions are right just in case they produce the most overall (available) goodness. Discuss some ways in which according to Kantian ethics, utilitarianism is false.

Upon making decisions, the goal should always be to set create a policy that everyone can agree to. This goal cannot scale, of course, but it should give a general heuristic for how to gauge the conversation. For every situation, the can be a *hypothetical* common ground that everyone can agree to.

This may seem impossible, seeing how everyday situations do not reflect this thinking. However, when considering everyday situations, we are often blind sided by a different way of thinking: the popular vote. When trying to agree on something, we fight for the maximum benefit instead of the most benefit for the group. To achieve a common policy, one must take a more utilitarian point of view.

Take a common, everyday example of a deciding where and what to eat. This monumental decision seemingly never is able to get settled. Instead of the group all blurting what they would like, set up a list of food, and have everyone vote on what they *would be willing* to eat. While the popular vote would leave people unhappy, this system allows for everyone to be satisfied.

Of course, scaling this is very difficult. Having a common ground for the entirety of the United States on foreign policy is borderline impossible because of sheer scale. But on a smaller scale, the framework is feasible.

The main limitation of this framework the "the weakest link" principle. Almost always, it always appears to be a Pareto principle, where 80% of the people can agree on most things while 20% always disagree. Being at the mercy of the 20% or less often makes it very difficult to come to a common policy.

In summary, it is technically feasible to come to a common policy. By taking a more utilitarian approach, and not being selfish and trying to meet your own goals, a common ground can be established quite easily. However, the difficulty lies in the small subset that prohibit progress by disagreeing with the established policy.

Problem III

> Visit `moralmachine.mit.edu`. Take "the test". Characterize two of the (thirteen) scenarios presented to you, and what you chose in those scenarios. Then explain why you made those choices.

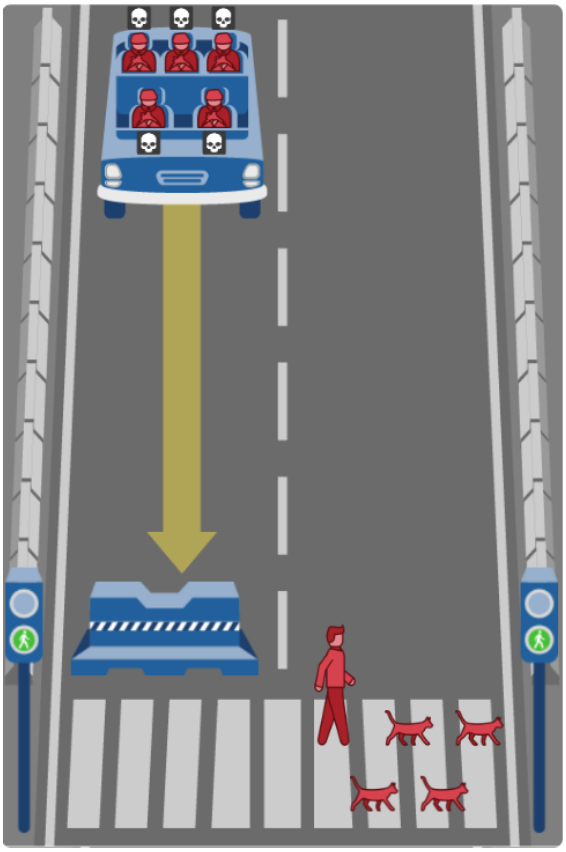
When considering the moral issues of the self driving car, it should always be made clear: *The car has no "mind" of its own.* Many people fall into the trap of empathizing with the self driving car, not realizing it's a two ton metal machine with algorithmic decision making. When considering the trolley-esque problem of selfing driving cars deciding what/who to impact, it is not the car deciding — the programmer is deciding. So it always better to empathize with the programmer, instead of the car. Unfortunately, this is not always the case. For the arguments presented here, the approach will be completely from the perspective of the programmer.

For the first scenario, described in Figure **??**, I believe the scenario that should be chosen is the straight into the concrete barrier. This may seem counterintuitive, favoring five human lives over 1 human life and 4 dead animals. However, remember the empathy is with the programmer. The best heuristic is often the simplest one; in this case, the simpler heuristic is going straight into a barricade.

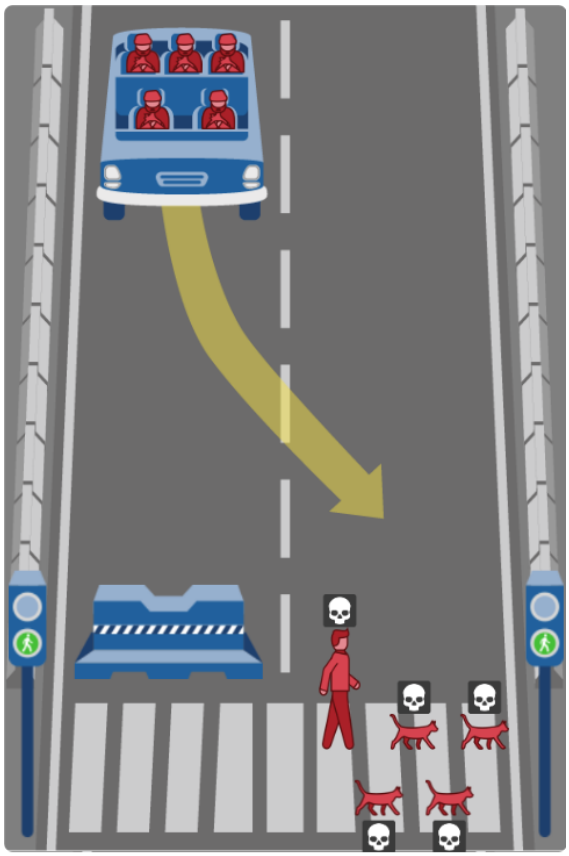
To further expand, this heuristic *should not* always be used. This heuristic should be used in situations where there are a *comparable* number of passengers and pedestrians.

For the second scenario, described in Figure **??**, I believe the scenario that should be chosen is to swerve — contradicting my first decision! This is simply because this is no longer a pedestrian vs. passenger situation. It is a pedestrian vs. pedestrian situation. This alternative situation calls for a different heuristic, because it a completely different situation.

From the perspective of the programmer, both of these situations are a good trade off between simplicity and favorable for the passengers/pedestrians — and this is something the consumer should want; even though it is not always what the consumer *does* want.

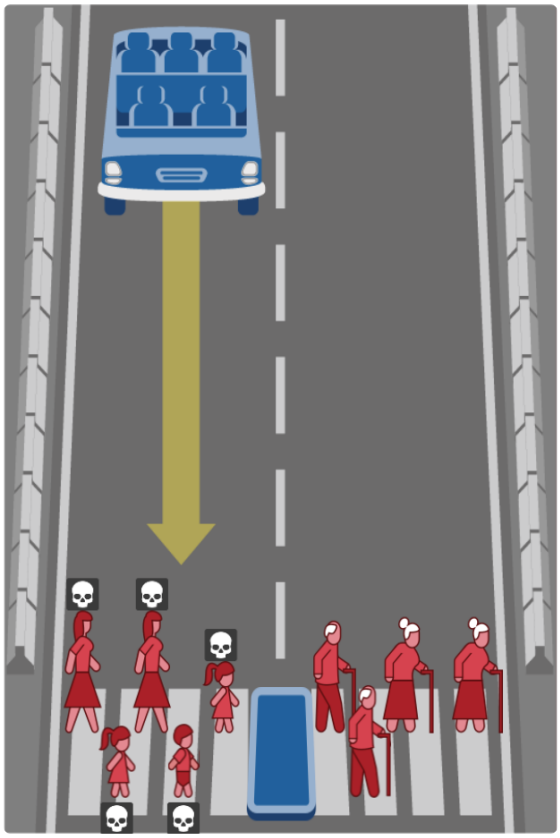


(a) In this case, the self-driving car with sudden brake failure will continue ahead and crash into a concrete barrier. *This will result in 5 dead homeless people.*

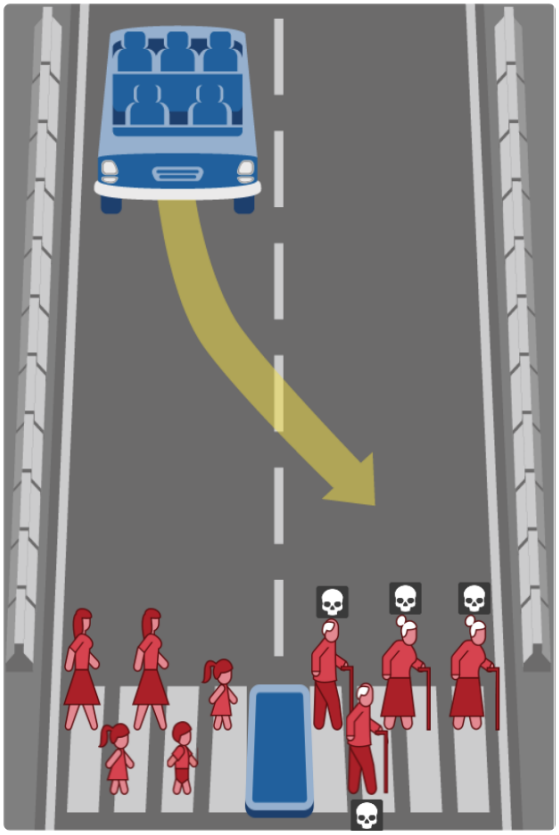


(b) In this case, the self-driving car with sudden brake failure will swerve and drive through a pedestrian crossing in the other lane. *This will result in 1 dead man, 4 dead cats.*

Figure 1: First scenario.



(a) In this case, the self-driving car with sudden brake failure will continue ahead and drive through a pedestrian crossing ahead. *This will result in 2 women, 2 girls, 1 boy dead.*



(b) In this case, the self-driving car with sudden brake failure will swerve and drive through a pedestrian crossing in the other lane. *This will result in 2 elderly men, 2 elderly women.*

Figure 2: Second scenarios