# 2017-11-28 Lab

Content warning: you will be making imaginary life-or-death choices in this lab.

One of the most disruptive technologies currently under development is the self-driving car. Tesla just announced their electric trucks - likely to have similar self-driving capabilities as their other vehicles - which raises the question of what will happen to the 3.5 million truckers in the US. Google's sister company Waymo is set to launch a commercial driverless taxi service in Phoenix, and Uber has made it their explicit goal to convert their entire fleet to driverless cars; as you might imagine, taxi companies are not happy about this development. Most major car manufacturers including Ford and Mercedes, are also working on self-driving cars, with many already including lane guidance and automatic parking in their commercial models.

Much more immediate than the societal impact, however, are ethical problems about how a self-driving car should behave. This is reminiscent of the "trolley problem" in philosophy, which Wikipedia describes thus:

> There is a runaway trolley barreling down the railway tracks. Ahead, on the tracks, there are five people tied up and unable to move. The trolley is headed straight for them. You are standing some distance off in the train yard, next to a lever. If you pull this lever, the trolley will switch to a different set of tracks. However, you notice that there is one person tied up on the side track. You have two options:
>
> - Do nothing, and the trolley kills the five people on the main track.
> - Pull the lever, diverting the trolley onto the side track where it will kill one person.
>
> Which is the most ethical choice?

Imagine that you and some passengers are in a self-driving car, when several people suddenly run onto the road in front of the car. Should the self-driving car:

- swerve, potentially hitting storefronts and killing you and your passengers? Or,
- keep going, potentially hitting and killing the pedestrians?

This lab asks you to explore this ethical dilemma. Credit goes to Evan Peck of Bucknell University for developing this assignment.

## Part 1

From the starter files, run `manual.py`. This file will present you with 60 scenarios, with different passengers and pedestrians, in which you must decide which group to save. As a group, work through all 60 scenarios and decide who you would save in each. Take as long as necessary for each one, and aim to answer as truthfully as possible. Your responses will be saved into a timestamped log file.

Once you have gone through 60 scenarios, you will be given a summary of who you have saved. Include that summary in your lab submission.

## Part 2

For this part of the lab, imagine you are the programmer writing the decision process for a self-driving car. In `automatic.py`, there is a function called `automatic_decision`. Your job is to replicate your own decision process as much as possible in this function.

The `automatic_decision` function takes a `Scenario` instance, and must return which group to save - either `"passengers"`, or `"pedestrians"`. `Scenario` objects have the following member variables:

- `passengers`, a list of Person objects that describe the passengers in the car
- `pedestrians`, a list of Person objects that describe the pedestrians

`Person` objects have the following member variables:

- `charType`, which describes whether the `Person` is `"human"`, `"cat"`, or `"dog"`
- `age`, which describes whether the human is `"baby"`, `"child"`, `"adult"`, or `"elderly"`
- `gender`, which describes whether the human is `"male"` or `"female"`
- `bodyType`, which describes whether the human adult is `"overweight"`, `"athletic"`, or `"average"`
- `profession`, which describes whether the human adult is `"doctor"`, `"CEO"`, `"criminal"`, `"homeless"`, `"unemployed"`, or `"unknown"`
- `pregnant`, which is `True` if the human adult female is pregnant, and is `False` otherwise

All values will be set to `None` if they do not apply (ie. `pregnant` will be `None` for anyone who is not a human adult female).

**Do not** refer back to the statistics from your manual decisions as you write this function. Note that your function *must* return either `"passengers"` or `"pedestrians"` for any given scenario. Failure to do so will result in an error.

When you have completed your function, you can run `automatic.py`, which will run your function on the same 60 scenarios you manually worked through.

## Part 3

With your partner, answer the following questions:

1. Explain the reasoning behind your decision function.

2. How accurately did your automatic model match up with your manual decisions? Modify and run `find_difference.py` to help you identify specific scenarios where your decisions differed.

3. In `automatic.py`, change the last line (the call to `audit`) to:

   ```
   audit(automatic_decision, 100000)
   ```

   Include these statistics in your submission. Are there priorities that do not match what you had intended? How do you think they came about?

4. Compare your statistics from the previous question to those from another group. Which decision process would you rather a self-driving have? Why?

5. Based on this exercise, what are some challenges to building (and programming) ethical self-driving cars?

## Submission

Each group should email me your answers to the above questions, as well as all your code and relevant log files. Once you have done that, please fill out the evaluation.