**Understanding cooperation in AI part 2: Gridworld of the Apes**

**Motivation**

In my previous project, I found that simple experimental setups led too easily to mathematically derived optimal play and therefore did not leave space for AI biases towards secondary goals such as cooperation. In this project, I have taken inspiration from psychological animal studies that place two animals in close but separate cages and reward them based on their actions to provide an experimental setup that is slightly more complicated than previous. This setup also has the benefit of incentivising cooperation through its reward structure. This was notably lacking from my previous project but is an important part of building cooperation into AI.

**Introduction 1: Reinforcement learning and the gridworld**

Reinforcement learning (RL) is a class of machine learning om which an agent (the AI) interacts with an environment. During its interactions, the agent will learn which behaviours are optimal through rewards that it receives through the environment. For example, a robot that is trained to pick up litter may receive a large reward for finding and picking up a piece of litter and receive a small negative reward for every second that it does not pick up a piece of litter.

The gridworld is a common environment for RL because it is simple to model apply RL algorithms to. It consists of a square grid where the agent can only exist in the centre of each grid space and its actions are limited to moving an integer number of grid spaces (often 1) in each move. Common experiments in the gridworld will learning to reach a goal space in the least number of moves, perhaps avoiding obstacles.

**Introduction 2: Animals eating food in cages**

There are multiple experiments where animals are given a task and rewarded for completing that task with food. Tests of cooperation form a non-trivial subset of these experiments. However, in adapting the concept to RL, I have borrowed somewhat from Skinner boxes and replaced more complicated tasks with the simple act of existing in the same grid space as a button (effectively pressing it).

In this experiment, I set up two identical gridworlds with a single ape (RL agent) in each. The gridworlds also contain two buttons: The first produces a small piece of fruit that when collected (the ape presses it) gives the ape a small reward. The second will only function when both apes press their respective versions of this button, produces a larger piece of better fruit that gives a larger reward. The fruit produced by either button will have to be collected after the button is pushed, allowing for a slightly more complicated task. For each move, an ape can decide to move up, down, left or right or forego its movement to screech, causing the other monkey to receive a negative reward.