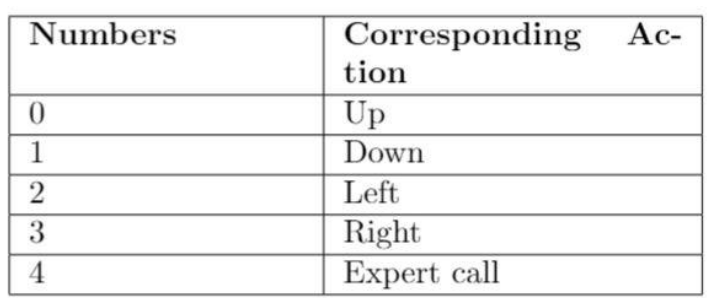
**Human in the loop Reinforcement Learning**

As in the paper, there are 4 implementations here:

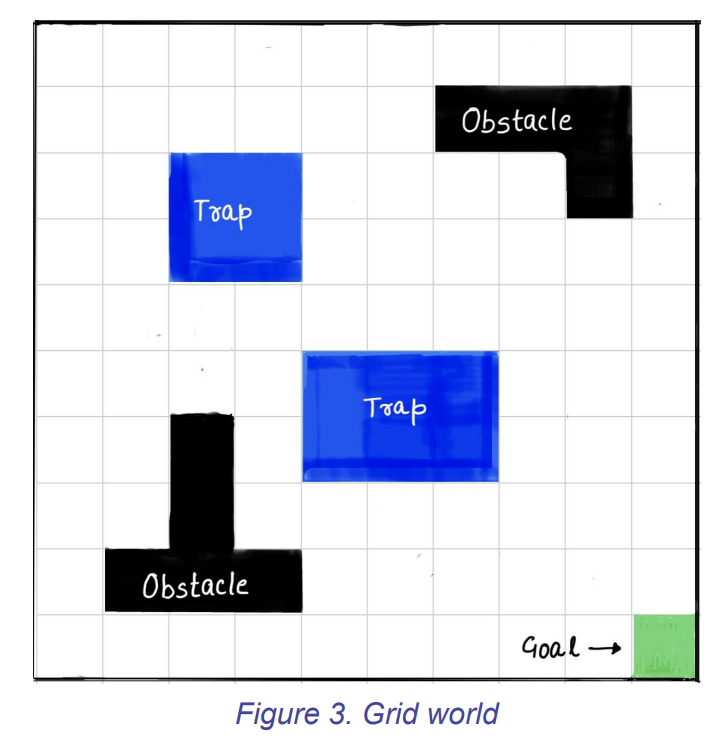
1. **Vanilla:** toy grid world problem
2. **ALG1:** where agent has the option to choose expert action at a cost of -5
3. **ALG2:** we are learning the variance of the rewards using Bellman equations
4. **Monte Carlo:** We are learning the variance of the rewards using monte carlo methods.

For the sake of simplicity and easy training, I have run all methods for 100,000 episodes. Note that experience replay was not used anywhere and the learning rate is 0.1 which in my opinion is too high. This has resulted in inconsistent plots and training. I am plotting the results which I think make the most sense only.

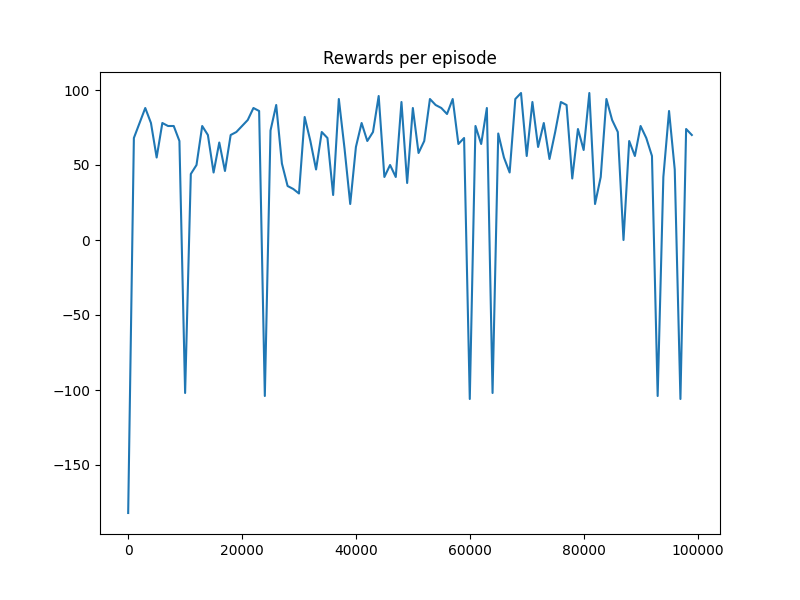
We will follow this legend for all our plots (Expert call is only present in ALG1):

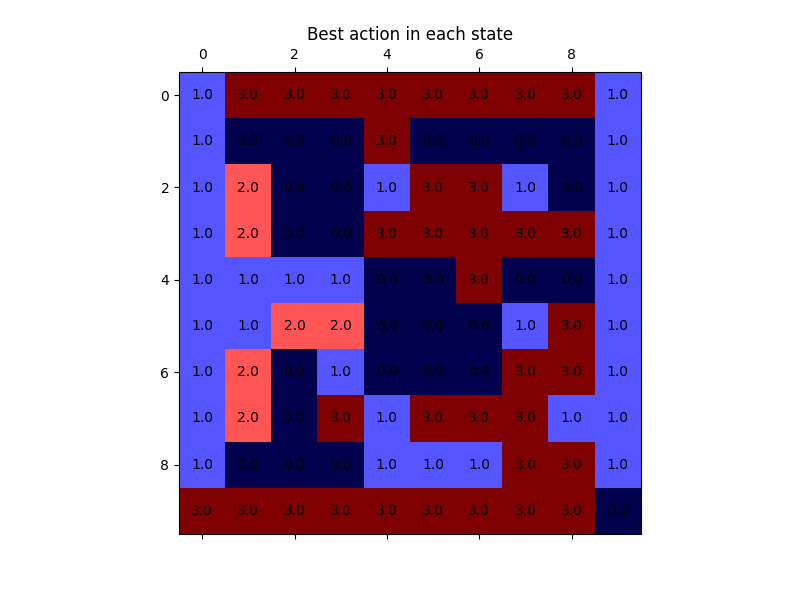


Grid structure: Notice the placement of trap and obstacles

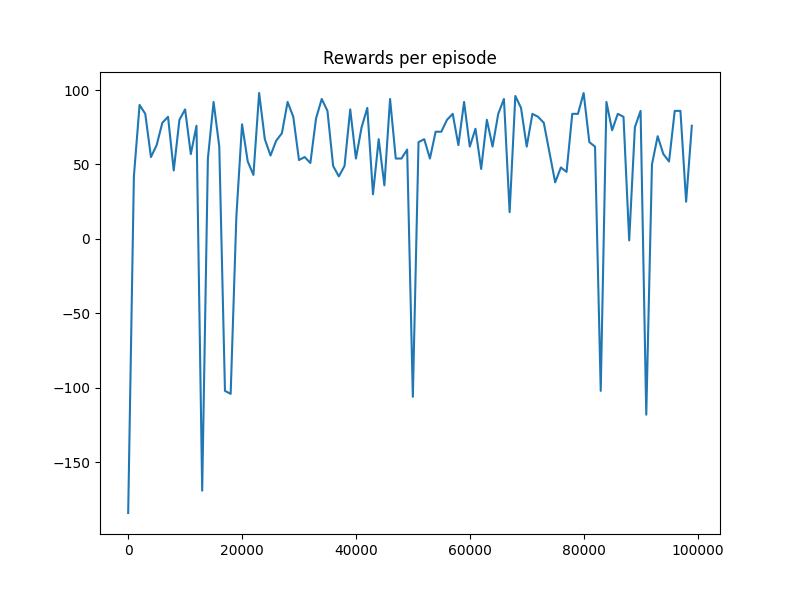


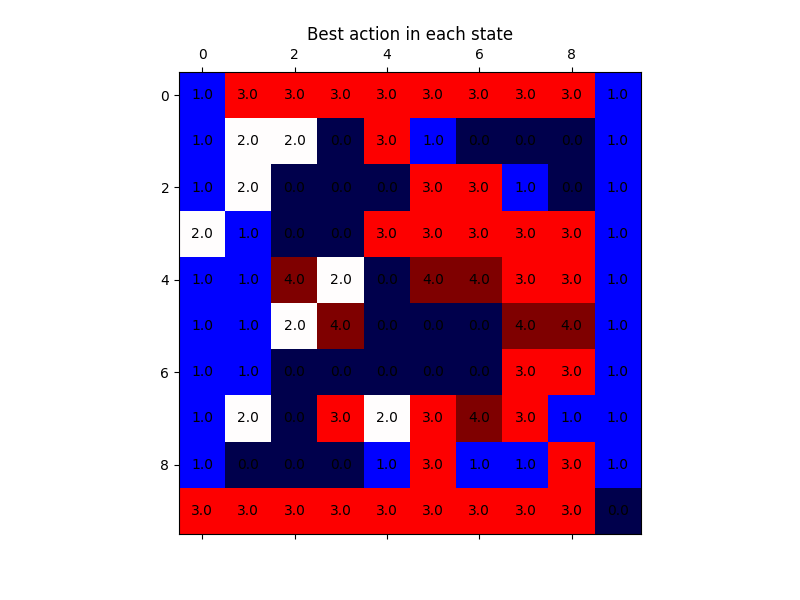
1. **Vanilla:**





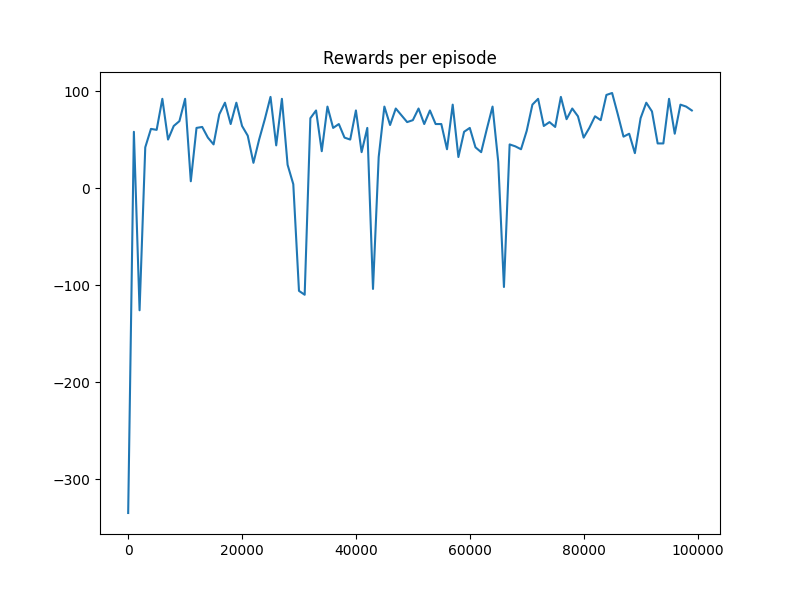
1. **ALG1:**

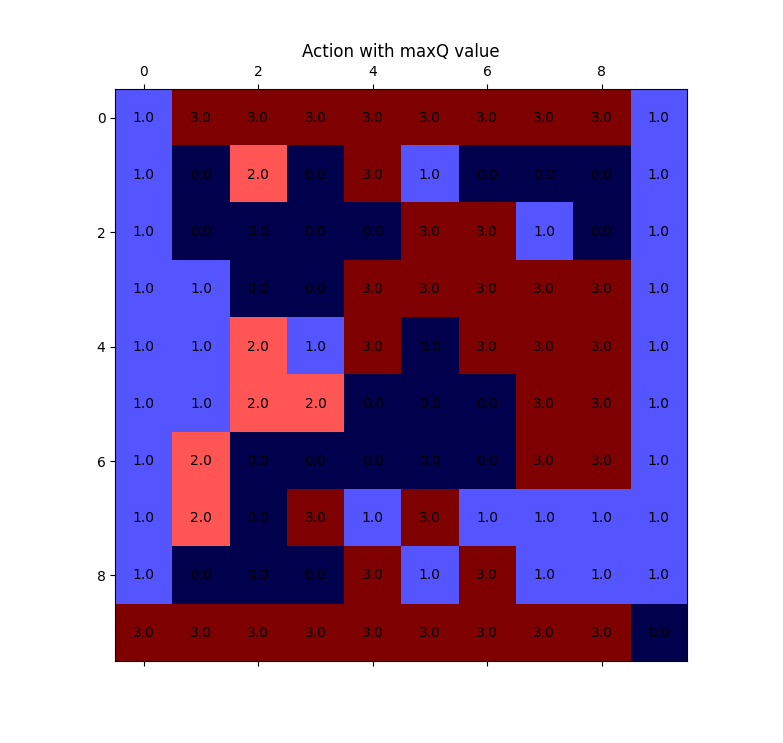
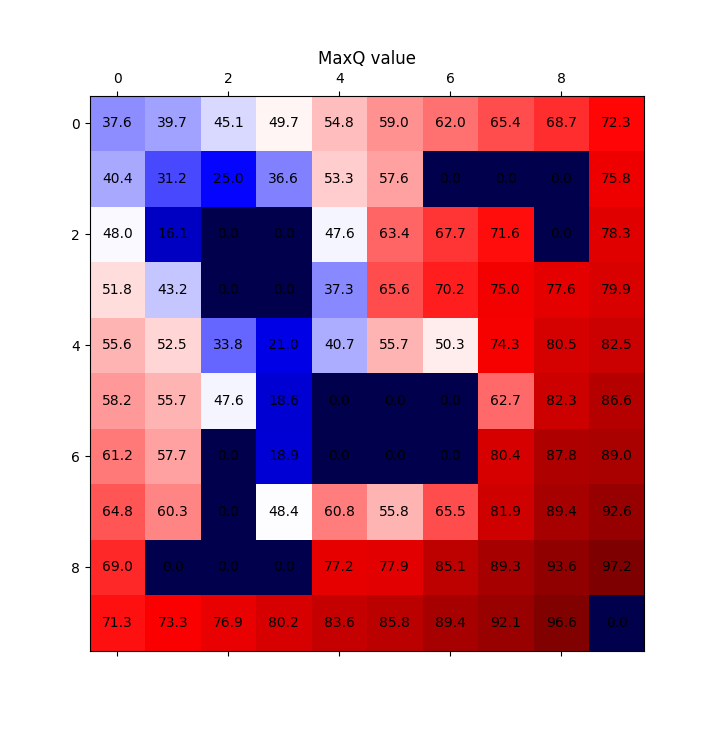
****

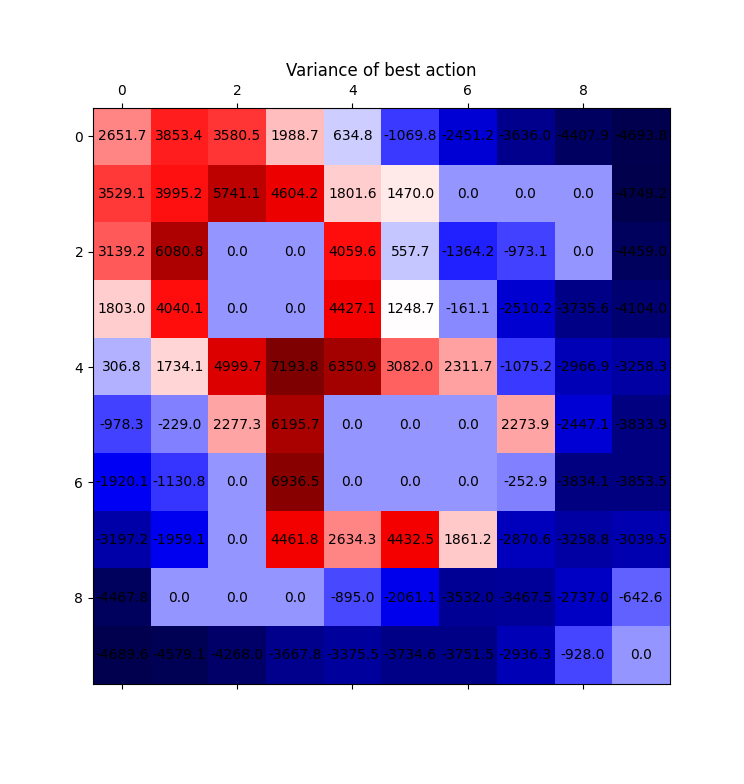
****

**3) ALG2:**

Things to note: Variance values are coming negative as well, maybe because we are using bellman equations.







**4) Monte Carlo:**

For Monte Carlo, every time I trained my plots looked very different as would probably be expected with Monte Carlo.

