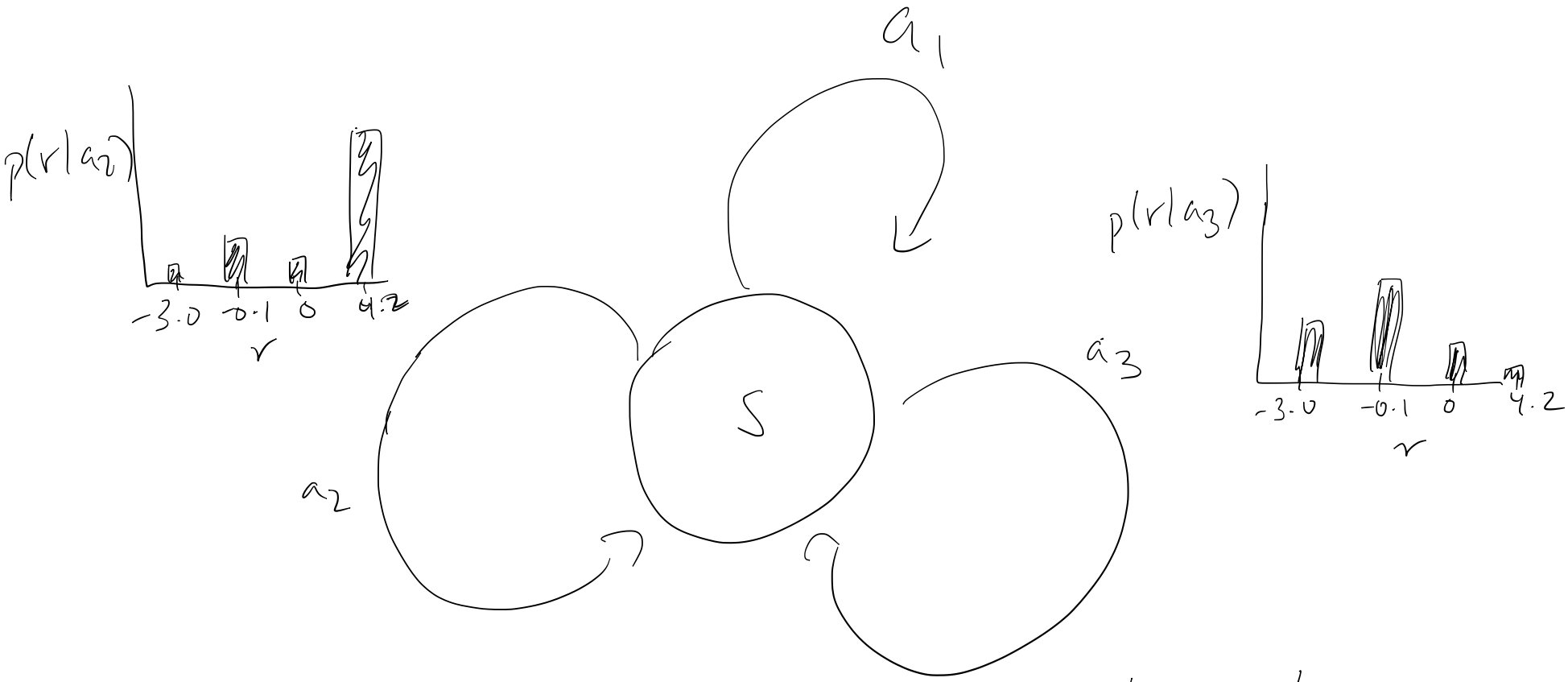
Bandit as MDP:



- Single state, agent always transitions back to s for every action
- Different distributions over rewards for each action