Reinforcement Learning: Chapter 3 Exercises

Name: Eli Andrew

(a) **Exercise 3.11:** If the current state is S_t, and actions are selected according to stochastic policy π, then what is the expecation of $R_t + 1$ in terms of π and the four-argument function p (3.2)?

- (3.2) states: $p(s', r|s, a) \doteq Pr S_t = s', R_t = r|S_{t-1} = s, A_{t-1} = a$
- This equation gives the probability of being in s' and receiving r given that you were previously in state s and took action a.
- Expected reward in the next time step R_{t+1} is equal to the rewards received from every action you can take from S_t multiplied by their probability of occuring.
- This gives R_{t+1} as the sum over all actions of the probability of taking the particular action multiplied by the reward from taking the action:
  $\sum_{a \in A} \pi(a|S_t) p(s', r|S_t, a)$