Reward Function

- Formally:
  - Define $R(s, a, s')$ for all $s, s' \in S$ and for all $a \in A$

Other possible ways to define reward functions

- Alternate forms: (given $R(s, a, s')$ above)
  - $R(s, a)$ as $\sum_{s'} P(s'|s, a)R(s, a, s')$ independent of s'
  - $R(s)$ as $\sum_{s'} P(s'|s, a)R(s, a, s')$, independent of a and s'
- Challenges
  - It is hard to construct reward functions with multiple attributes
  - Balance risk vs reward

Utility of State and Optimal Policy

- Utility of state (or value of state) is the value of optimal policy

$$U(s) = U^{\pi^*}(s) = \max_{a \in A(s)} \sum_{s'} P(s'|s,a)[R(s,a,s') + \gamma U(s')] = V(s)$$

  - Expected sum of discounted rewards if an optimal policy is executed
    - $R(s) = \sum_{s'} P(s'|s,a,)R(s,a,s')$ is the "short term" reward for being in s
    - $U(s) = V(s)$ is the "long term" total expected reward from s onward
- Optimal action selected through maximizing utility of state $U(s)$ based on MEU:

$$\pi^*(s) = \text{argmax}_{a \in A(s)} \sum_{s'} P(s'|s,a)[R(s,a,s') + \gamma U(s')]$$

  - Optimal policy independent of starting state in infinite horizon problems with discounted utilities