

Recap: MDPs

- Markov decision processes:
 - States S
 - Actions A
 - Transitions $P(s' | s, a)$ (or $T(s, a, s')$)
 - Rewards $R(s, a, s')$ (and discount γ)
 - Start state s_0

- Quantities:
 - Policy = map of states to actions
 - Utility = sum of discounted rewards
 - Values = expected future utility from a state (max node)
 - Q-Values = expected future utility from a q-state (chance node)

