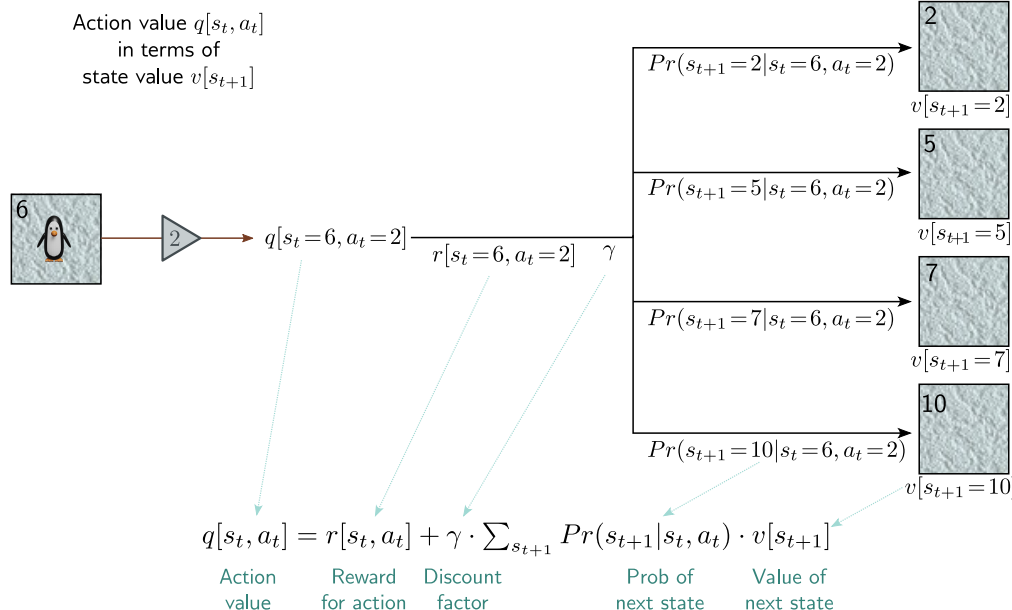
Action value $q[s_t, a_t]$
in terms of
state value $v[s_{t+1}]$
6
2
$q[s_t=6, a_t=2]$
$r[s_t=6, a_t=2]$
γ
$Pr(s_{t+1}=2|s_t=6, a_t=2)$
$Pr(s_{t+1}=5|s_t=6, a_t=2)$
$Pr(s_{t+1}=7|s_t=6, a_t=2)$
$Pr(s_{t+1}=10|s_t=6, a_t=2)$
2
5
7
10
$v[s_{t+1}=2]$
$v[s_{t+1}=5]$
$v[s_{t+1}=7]$
$v[s_{t+1}=10]$
$q[s_t, a_t] = r[s_t, a_t] + \gamma \cdot \sum_{s_{t+1}} Pr(s_{t+1}|s_t, a_t) \cdot v[s_{t+1}]$
Action value
Reward for action
Discount factor
Prob of next state
Value of next state