3.17

$$q_\pi(s,a) = E_\pi[G_t \mid S_t = s, A_t = a] \quad \text{\# by def}$$

$$= E_\pi\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \mid S_t = s, A_t = a\right]$$

$$= E_\pi\left[R_{t+1} + \gamma \sum_{k=0}^{\infty} \gamma^k R_{t+k+2} \mid S_t = s, A_t = a\right]$$

$$= \sum_{s' \in S} p(s' \mid s,a) \cdot r(s,a,s') \quad \text{\# by defn of expectation and linearity of expectation}$$

$$+ \sum_{s' \in S} p(s' \mid s,a) \cdot E_\pi\left[\gamma \sum_{k=0}^{\infty} \gamma^k R_{t+k+2} \mid S_t = s, A_t = a, S_{t+1} = s'\right]$$

$$\gamma \sum_{a' \in A} \pi(a' \mid s') \cdot E_\pi\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+2} \mid S_t = s, A_t = a, S_{t+1} = s', A_{t+1} = a'\right]$$

\# defn of expectation

$$\gamma \sum_{a' \in A} \pi(a' \mid s') \cdot E_\pi\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+2} \mid S_{t+1} = s', A_{t+1} = a'\right]$$

\# By Markov

$$E_\pi[G_{t+1} \mid S_{t+1} = s', A_{t+1} = a']$$

\# By defn of G_t

This gives us

$$\sum_{s' \in S} p(s' \mid s,a) \cdot r(s,a,s') +$$

$$\sum_{s' \in S} p(s' \mid s,a) \cdot \gamma \sum_{a' \in A} \pi(a' \mid s') \cdot q_\pi(s', a')$$

\# Substituting ① and by defn of $q_\pi(s,a)$

$$q_\pi(s,a) = \sum_{s' \in S} p(s|a,s') \cdot [r(s,a,s') + \gamma \sum_{a' \in A} \pi(a'|s') q_\pi(s',a')]$$

* By linearity of expectation

3.19

$$v_\pi(s) = \sum_{a \in A} \pi(a|s) q_\pi(s,a)$$

Note from above

② $$q_\pi(s,a) = \sum_{s' \in S} p(s|a\,s') \cdot [r(s,a,s') + \gamma v_\pi(s')]$$

$$q_\pi(s,a) = E[R_{t+1} | S_t = s, A_t = a] + \gamma E[v_\pi(S_{t+1}) | S_t = s, A_t = a]$$

* from above derivation in q 3.17 just backward

$$q_\pi(s,a) = E[R_{t+1} + \gamma v_\pi(S_{t+1}) | S_t = s, A_t = a]$$

Now we simply apply (3.4) to ②

3.4 $$p(s'|s,a) = \sum_{r \in R} p(s', r | s, a)$$

$$q_\pi(s,a) = \sum_{s \in S} \sum_{s' \in S} \sum_{r \in R} p(s', r | s, a) \cdot [r + \gamma v_\pi(s')]$$