Exercise 3.4

s	a	s'	$p(s', r \mid s, a)$	r
high	wait	high	1.	r_{wait}
high	search	high	α	r_{search}
high	search	low	$1-\alpha$	r_{search}
low	wait	low	1	r_{wait}
low	search	high	$1-\beta$	-3
low	search	low	β	r_{search}
low	recharge	high	1	0.

Exercise 3.15

$$G = R_1 + \gamma R_2 + \gamma^2 R_3 + \cdots + \gamma^T R_T$$

$$G' = (R_1 + C) + \gamma(R_2 + C) + \cdots + \gamma^T (R_T + C)$$

$$= (R_1 + \gamma R_2 + \cdots \gamma^T R_T) + C + \gamma C + \gamma^2 C + \cdots + \gamma^T C$$

$$= G + \frac{C(1-\gamma^T)}{1-\gamma} \quad \text{as } T \to \infty$$

$$= G + \frac{C}{1-\gamma}$$ hence it doesn't affect the relative ordering.

$$V_C > \frac{C}{1-\gamma}$$

Exercise 3.16

Continuing the previous question, we got

$$G' = G + \frac{c(1-\gamma^T)}{1-\gamma}$$

Now how long the episode will last will also play a role in comparing.

The longer the episode last, the more reward we will get, as $(1-\gamma^T)$ increases with increase in T as $0 < \gamma < 1$.

$$G'_{T=1} = G + \frac{c(1-\gamma)}{1-\gamma}$$

$$= G + c$$

$$G'_{T=2} = G + \frac{c \cdot (1-\gamma^2)}{1-\gamma} = G + c(1+\gamma)$$

$$G'_{T=2} > G'_{T=1}.$$

Q5 Ans:

$$V_*(s) = \max_{a \in A(s)} q_{\pi_*}(s,a).$$