APOORV KHATTAR
2016016

Ques 1.

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

↖ Exercise 3.4

Ques 5. $V_*(s) = \max_{a \in A(s)} q_{\pi^*}(a, s)$

$= \max_a E_{\pi^*}[G_t \mid A_t = a, S_t = s]$

$= \max_a E_{\pi^*}[R_t + \gamma G_{t+1} \mid A_t = a, S_t = s]$

$= \max_a E_{\pi^*}[R_t + \gamma V_*(S_{t+1}) \mid A_t = a, S_t = s]$

$= \max_a \sum_{s'} \sum_r p(s', r \mid s, a) \{r + \gamma V_*(s')\}$

$= \max_a \sum_{s'} \sum_r p(s', r \mid s, a) \{r + \gamma \max_a q_*(s', a)\}$

Ques 3.

(a) Exercise 3.15

$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} \ldots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$ (1)

If we add constant to all rewards then,

$G_t' = (R_{t+1} + c) + \gamma (R_{t+2} + c) \ldots = \sum_{k=0}^{\infty} \gamma^k (R_{t+k+1} + c)$

$= \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} + \sum_{k=0}^{\infty} \gamma^k c$

$\Rightarrow G_t' = G_t + c\left(\frac{1}{1-\gamma}\right)$

$\Rightarrow G_t' = G_t + V_c$

Let $V_\pi'(s)$ be the new value function for G_t' then

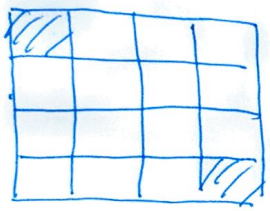
$$V_\pi'(s) = E_\pi[G_t' \mid S_t = s]$$
$$= E_\pi[G_t + V_c \mid S_t = s]$$
$$= E_\pi[G_t \mid S_t = s] + E_\pi[V_c \mid S_t = s]$$
$$= V_\pi(s) + V_c$$

$\Rightarrow V_\pi'(s)$ increases by V_c and $V_c = \frac{c}{1-\gamma}$

(b) Exercise 3.16

Adding a constant (c) to all rewards in an episodic task will effect the tasks.

Consider the following episodic task:-



(0,0) and (3,3) are terminal states

reward for each action (-1)

Optimal Policy would ensure shortest path to either of the terminal states.

Now if the rewards are $-1 + c$ $(c > 1)$ then optimal policy changes, also value functions will for any policy will be different as it would never try to go to terminal states for maximizing reward