A3) $R = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} \ldots\ldots$

$\therefore C \rightarrow$ Constant.

$R^C = (R_{t+1} + C) + \gamma(R_{t+2} + C) \ldots\ldots$

Then $V_\pi(s) = E[R|s]$

& $\tilde{V}_\pi(s) = E[R^C|s]$

$$= E\left[(R_{t+1}+C) + \gamma(R_{t+2}+C) + \gamma^2(R_{t+3}+C)\middle|s\right]$$

$$= E\left[\sum_{k=1}^{\infty} \gamma^{k-1} R_{t+k}\middle|s\right] + E\left[C\sum_{k=0}^{\infty}\gamma^k\middle|s\right] \quad \text{where } 0 < \gamma < 1$$

($\uparrow$ constant)

$$= E[R|s] + C\sum_{k=0}^{\infty}\gamma^k$$

$$\boxed{\tilde{V}_\pi(s) = V_\pi(s) + \frac{C}{1-\gamma}}$$

$\therefore \sum_{k=0}^{\infty}\gamma^k = \frac{1}{1-\gamma}$

So relative evaluation does not change.

for episodic Task,

$$\tilde{V}_\pi(s) = V_\pi(s) + C\sum_{k=0}^{T}\gamma^T$$

$$\boxed{\tilde{V}_\pi(s) = V_\pi(s) + \frac{C(1-\gamma^T)}{1-\gamma}}$$

Hence as we go closer to terminal state the change in $\tilde{V}_\pi(s)$ increases

A5) $V_*(s) = \max_\pi V(s) \quad \forall s$ —① $\quad q_*(s) = \max_\pi q_\pi(s,a)$ —②

$\Rightarrow$

$V_\pi(s) = q_\pi(s, \pi(s))$.

~~$V_*(s) = \max_\pi q_\pi(s, \pi(s))$~~ $V_*(s) = \max_\pi q_\pi(s, \pi(s))$.

~~$V_*(s) = \max_a$~~ $V_*(s) = \max_a$ as optimal $\pi(s) = a$.

So $V_*(s) = \max_a \max_\pi q_\pi(s,a)$

$$\boxed{V_*(s) = \max_a q_*(s,a)}$$

considered them to be constants from function T.

A1)

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

* As for each tuple (s, a, s'), only 1 reward exists

$$\sum_{r} p(s', r \mid s, a) = p(s' \mid a, s) \quad * \qquad (s, a, s') \rightarrow T$$

so $$p(s', T(s, a, s') \mid s, a) = p(s' \mid a, s)$$