Exercise 3.15

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

$$G' = (R_1 + c) + \gamma (R_2 + c) + \dots + \gamma^T (R_T + c)$$

$$= (R_1 + \gamma R_2 + \dots \gamma^T R_T) + c + \gamma c + \gamma^2 c + \dots + \gamma^T c$$

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

$$= G + \frac{c}{1-\gamma}$$ hence ~~it is the~~ 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 $(1-\gamma^T)$ ~~decreases~~ increases with increase in T as $0<\gamma<1$.

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

$$= G + c$$

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

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

Q_5 Ans:

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

Q1. Since $r(s, a s')$ is an expectation we need to figure out the ~~actual reward~~ ~~for each~~ probability of finding a can or not finding a can.

$p \leftarrow$ find a can $q \leftarrow$

$r_{search} = p \times 1 + (1-p) \times 0$

$= p$.

$\therefore$ probability of finding a can is $\beta \cdot r_{search}$
~~Not~~ " " not " " " " $1 - r_{search}$

Given $p(s', r \mid s, a) = \alpha \times r_{search}$.

~~$p(s', r \mid s = \text{find\_can}, a) = \alpha r_{search}$~~
$p($~~$s', r \mid s = \text{not\_find}, a$~~

$p(s' \mid s, a) = \alpha$ so,

$p(s', r=0 \mid s, a) + p(s', r=1 \mid s, a) = p(s' \mid s, a)$

$\Rightarrow p(s', r=0 \mid s, a) = \alpha - \alpha r_{search}$.

Doing this for others.

s	a	s'	r	p(s', r | s, a)
H	S	h	1	$\alpha\, r_{search}$
h	S	h	0	$\alpha - \alpha\, r_{search}$
h	S	l	1	$(1-\alpha)\, r_{search}$
l	S	h	-3	$1-\beta$
l	S	l	1	$\beta \times r_{search}$
l	S	l	0	$\beta - \beta \times r_{search}$
h	S	l	1	$(1-\alpha) - ((1-\alpha) r_{search})$
h	W	h	1	r_{wait}
l	w	l	1	~~βr_{search}~~ r_{wait}
l	w	l	0	~~$\beta - \beta \times r_{wait}$~~ $1-r_{wait}$
l	r	h	0	1
h	W	h	0	$1-r_{wait}$