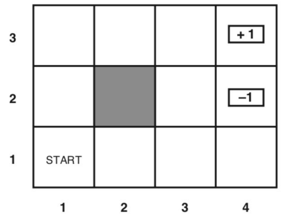
CS4246 / CS5446

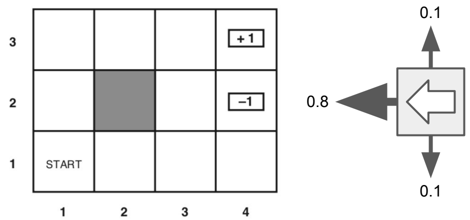
Tutorial Week 11

Muhammad **Rizki** Maulana
rizki@u.nus.edu

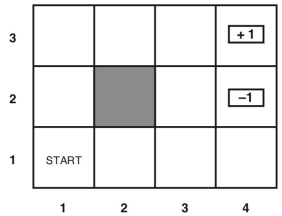
First



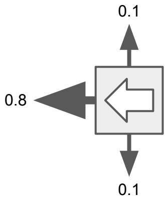
(a)



(a)



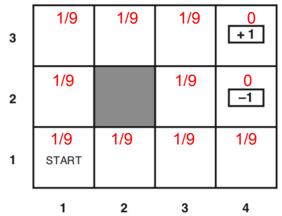
(a)



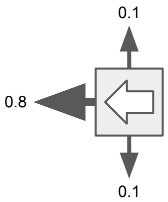
Detect adjacent wall

Noisy sensor:

- Correct : 0.9
- Wrong : 0.1



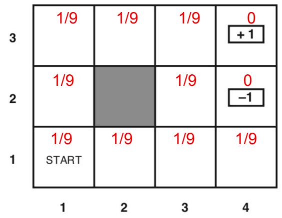
(a)



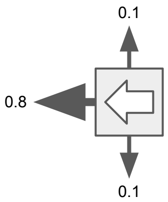
Detect adjacent wall

Noisy sensor:

- Correct : 0.9
- Wrong : 0.1



(a)



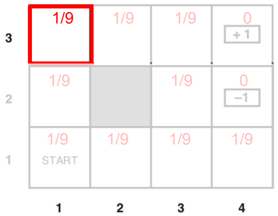
Calculate the exact belief state b_1 (rounded off to 5 decimal places) after the agent moves *Left* and its sensor reports 1 adjacent wall.

Detect adjacent wall

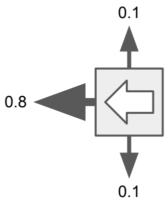
Noisy sensor:

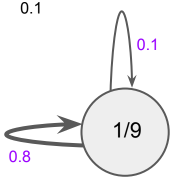
- Correct : 0.9
- Wrong : 0.1

Question



(a)



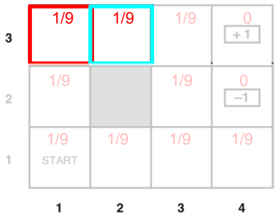


Detect adjacent wall

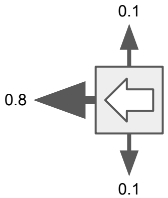
Noisy sensor:

- Correct : 0.9
- Wrong : 0.1

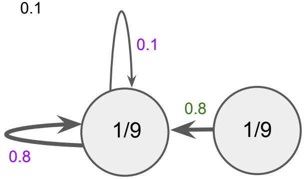
Calculate the exact belief state b_1 (rounded off to 5 decimal places) after the agent moves *Left* and its sensor reports 1 adjacent wall.



(a)



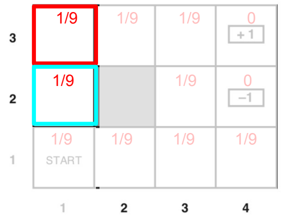
Calculate the exact belief state b_1 (rounded off to 5 decimal places) after the agent moves *Left* and its sensor reports 1 adjacent wall.



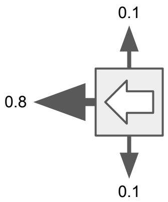
Detect adjacent wall

Noisy sensor:

- Correct : 0.9
- Wrong : 0.1



(a)

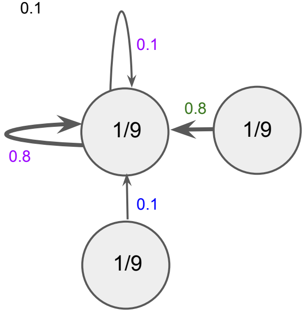


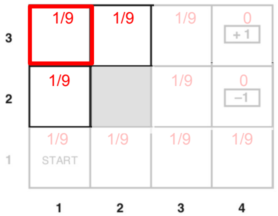
Calculate the exact belief state b_1 (rounded off to 5 decimal places) after the agent moves *Left* and its sensor reports 1 adjacent wall.

Detect adjacent wall

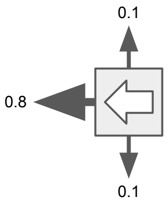
Noisy sensor:

- Correct : 0.9
- Wrong : 0.1





(a)

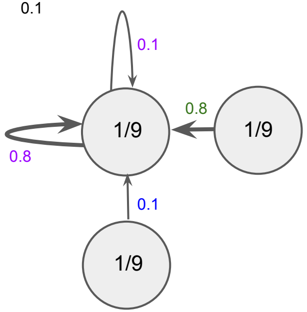


Calculate the exact belief state b_1 (rounded off to 5 decimal places) after the agent moves *Left* and its sensor reports 1 adjacent wall.

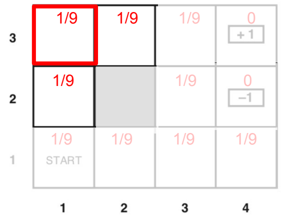
Detect adjacent wall

Noisy sensor:

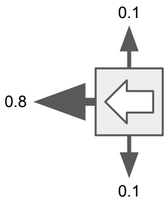
- Correct : 0.9
- Wrong : 0.1



$$P(x'|Left, b_0) = \sum_x P(x'|Left, x) b_0(x)$$



(a)

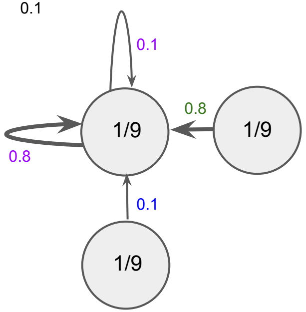


Calculate the exact belief state b_1 (rounded off to 5 decimal places) after the agent moves *Left* and its sensor reports 1 adjacent wall.

Detect adjacent wall

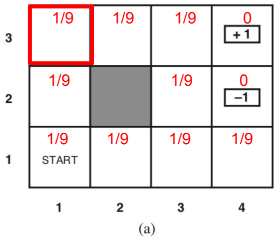
Noisy sensor:

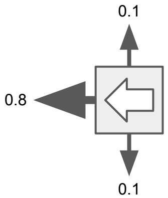
- Correct : 0.9
- Wrong : 0.1



$$P(x'|Left, b_0) = \sum_x P(x'|Left, x) b_0(x)$$

$$0.9(1/9) + 0.8(1/9) + 0.1(1/9) = 0.2$$



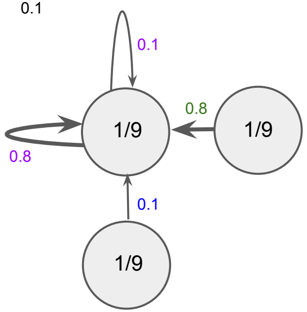


Calculate the exact belief state b_1 (rounded off to 5 decimal places) after the agent moves *Left* and its sensor reports 1 adjacent wall.

Detect adjacent wall

Noisy sensor:

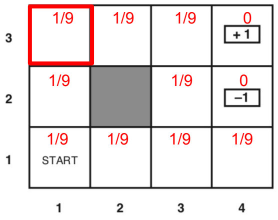
- Correct : 0.9
- Wrong : 0.1



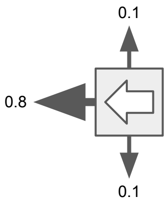
$$P(x'|Left, b_0) = \sum_x P(x'|Left, x) b_0(x)$$

0.9(1/9) + 0.8(1/9) + 0.1(1/9) = 0.2

0.2	$\frac{1}{9}$	$\frac{0.2}{9}$	0
$\frac{1}{9}$	×	$\frac{1}{9}$	$\frac{0.1}{9}$
0.2	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{0.1}{9}$



(a)

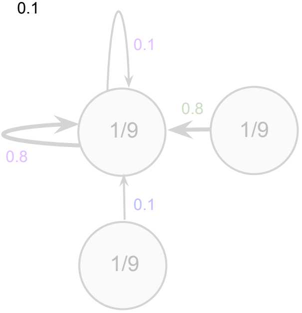


Calculate the exact belief state b_1 (rounded off to 5 decimal places) after the agent moves *Left* and its sensor reports 1 adjacent wall.

Detect adjacent wall

Noisy sensor:

- Correct : 0.9
- Wrong : 0.1

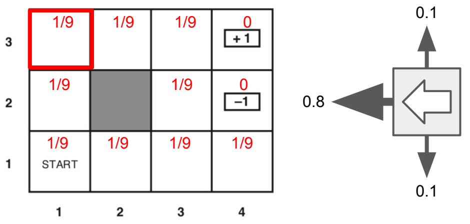


$$P(x'|Left, b_0) = \sum_x P(x'|Left, x)b_0(x)$$

0.9(1/9) + 0.8(1/9) + 0.1(1/9) = 0.2

0.2	$\frac{1}{9}$	$\frac{0.2}{9}$	0
$\frac{1}{9}$	×	$\frac{1}{9}$	$\frac{0.1}{9}$
0.2	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{0.1}{9}$

Now, we update these estimates with the sensor data, which says there is one adjacent wall (i.e., multiply by $P(z = \text{‘1 adjacent wall’}|x')$):



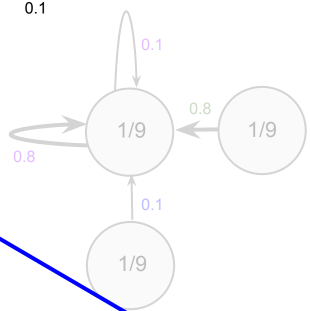
(a)

Calculate the exact belief state b_1 (rounded off to 5 decimal places) after the agent moves *Left* and its sensor reports 1 adjacent wall.

Detect adjacent wall

Noisy sensor:

- Correct : 0.9
- Wrong : 0.1



$$P(x'|Left, b_0) = \sum_x P(x'|Left, x) b_0(x)$$

0.9(1/9) + 0.8(1/9) + 0.1(1/9) = 0.2

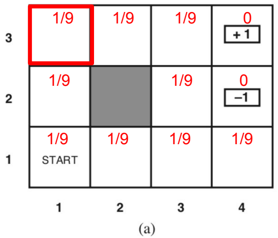
$\boxed{0.2}$	$\frac{1}{9}$	$\frac{0.2}{9}$	0
$\frac{1}{9}$	$\times$	$\frac{1}{9}$	$\frac{0.1}{9}$
0.2	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{0.1}{9}$

Now, we update these estimates with the sensor data, which says there is one adjacent wall (i.e., multiply by $P(z = \text{‘1 adjacent wall’}|x')$):

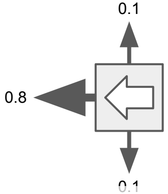
0.1×0.2	$0.1 \times \frac{1}{9}$	$0.9 \times \frac{0.2}{9}$	0
$0.1 \times \frac{1}{9}$	$\times$	$0.9 \times \frac{1}{9}$	$0.9 \times \frac{0.1}{9}$
0.1×0.2	$0.1 \times \frac{1}{9}$	$0.9 \times \frac{1}{9}$	$0.1 \times \frac{0.1}{9}$

1 adj wall

2 adj wall



(a)

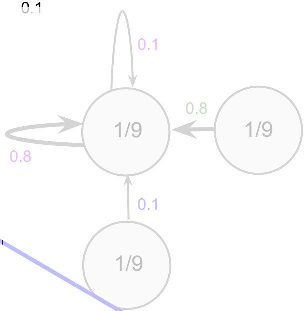


Calculate the exact belief state b_1 (rounded off to 5 decimal places) after the agent moves *Left* and its sensor reports 1 adjacent wall.

Detect adjacent wall

Noisy sensor:

- Correct : 0.9
- Wrong : 0.1



$$P(x'|Left, b_0) = \sum_x P(x'|Left, x) b_0(x)$$

0.9(1/9) + 0.8(1/9) + 0.1(1/9) = 0.2

0.2	$\frac{1}{9}$	$\frac{0.2}{9}$	0
$\frac{1}{9}$	×	$\frac{1}{9}$	$\frac{0.1}{9}$
0.2	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{0.1}{9}$

Now, we update these estimates with the sensor data, which says there is one adjacent wall (i.e., multiply by $P(z = \text{'1 adjacent wall'}|x')$):

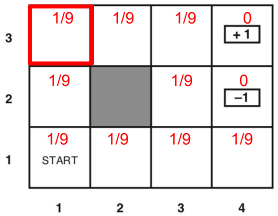
0.1×0.2	$0.1 \times \frac{1}{9}$	$0.9 \times \frac{0.2}{9}$	0
$0.1 \times \frac{1}{9}$	×	$0.9 \times \frac{1}{9}$	$0.9 \times \frac{0.1}{9}$
0.1×0.2	$0.1 \times \frac{1}{9}$	$0.9 \times \frac{1}{9}$	$0.1 \times \frac{0.1}{9}$

1 adj wall

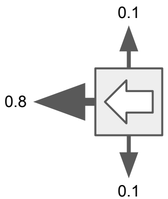
2 adj wall

and renormalize to get b_1:

0.06569	0.03650	0.06569	0
0.03650	×	0.32847	0.03285
0.06569	0.03650	0.32847	0.00365



(a)

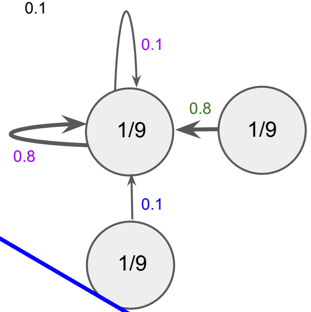


Calculate the exact belief state b_1 (rounded off to 5 decimal places) after the agent moves *Left* and its sensor reports 1 adjacent wall.

Detect adjacent wall

Noisy sensor:

- Correct : 0.9
- Wrong : 0.1



$$P(x'|Left, b_0) = \sum_x P(x'|Left, x) b_0(x)$$

0.9(1/9) + 0.8(1/9) + 0.1(1/9) = 0.2

$\boxed{0.2}$	$\frac{1}{9}$	$\frac{0.2}{9}$	0
$\frac{1}{9}$	$\times$	$\frac{1}{9}$	$\frac{0.1}{9}$
0.2	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{0.1}{9}$

Now, we update these estimates with the sensor data, which says there is one adjacent wall (i.e., multiply by $P(z = \text{'1 adjacent wall'} | x')$):

0.1×0.2	$0.1 \times \frac{1}{9}$	$0.9 \times \frac{0.2}{9}$	0
$0.1 \times \frac{1}{9}$	$\times$	$0.9 \times \frac{1}{9}$	$0.9 \times \frac{0.1}{9}$
0.1×0.2	$0.1 \times \frac{1}{9}$	$0.9 \times \frac{1}{9}$	$0.1 \times \frac{0.1}{9}$

1 adj wall

2 adj wall

and renormalize to get b_1:

0.06569	0.03650	0.06569	0
0.03650	$\times$	0.32847	0.03285
0.06569	0.03650	0.32847	0.00365

Second

[Modified from RN 3e 17.14] What is the time complexity of d steps of POMDP value iteration for a sensorless environment? Give an upper bound on the number of α-vectors generated in the process.

Question

[Modified from RN 3e 17.14] What is the time complexity of d steps of POMDP value iteration for a sensorless environment? Give an upper bound on the number of α-vectors generated in the process.

$$\alpha_p(s) = \sum_{s'} P(s'|s,a)[R(s,a,s') + \gamma \sum_{e'} P(e'|s')\alpha_{p.e'}(s')]$$

[Modified from RN 3e 17.14] What is the time complexity of d steps of POMDP value iteration for a sensorless environment? Give an upper bound on the number of α-vectors generated in the process.

sensorless

$$\alpha_p(s) = \sum_{s'} P(s'|s,a)[R(s,a,s') + \gamma \cancel{\sum_{e'} P(e'|s')}\alpha_{p.e'}(s')]$$

$$\alpha_p(s) = \sum_{s'} P(s'|s,a)[R(s,a,s') + \gamma \alpha_{p'}(s')]$$

[Modified from RN 3e 17.14] What is the time complexity of d steps of POMDP value iteration for a sensorless environment? Give an upper bound on the number of α-vectors generated in the process.

sensorless

$$\alpha_p(s) = \sum_{s'} P(s'|s,a)[R(s,a,s') + \gamma \cancel{\sum_{e'} P(e'|s')}\alpha_{p.e'}(s')]$$

$$\alpha_p(s) = \sum_{s'} P(s'|s,a)[R(s,a,s') + \gamma \alpha_{p'}(s')]$$

p = [a, p'], p' subplan

[Modified from RN 3e 17.14] What is the time complexity of d steps of POMDP value iteration for a sensorless environment? Give an upper bound on the number of α-vectors generated in the process.

sensorless

$$\alpha_p(s) = \sum_{s'} P(s'|s,a)[R(s,a,s') + \gamma \cancel{\sum_{e'} P(e'|s')\alpha_{p.e'}(s')}]$$

$$\alpha_p(s) = \sum_{s'} P(s'|s,a)[R(s,a,s') + \gamma \alpha_{p'}(s')]$$

p = [a, p'], p' subplan

$$V(b) = \max_p b \cdot \alpha_p$$

[Modified from RN 3e 17.14] What is the time complexity of d steps of POMDP value iteration for a sensorless environment? Give an upper bound on the number of α-vectors generated in the process.

sensorless

$$\alpha_p(s) = \sum_{s'} P(s'|s,a)[R(s,a,s') + \gamma \cancel{\sum_{e'} P(e'|s')}\alpha_{p.e'}(s')]$$

$$\alpha_p(s) = \sum_{s'} P(s'|s,a)[R(s,a,s') + \gamma \alpha_{p'}(s')]$$

p = [a, p'], p' subplan

$$V(b) = \max_p b \cdot \alpha_p$$

conditional plans

[Modified from RN 3e 17.14] What is the time complexity of d steps of POMDP value iteration for a sensorless environment? Give an upper bound on the number of α-vectors generated in the process.

sensorless

$$\alpha_p(s) = \sum_{s'} P(s'|s,a)[R(s,a,s') + \gamma \cancel{\sum_{e'} P(e'|s')}\alpha_{p.e'}(s')]$$

$$\alpha_p(s) = \sum_{s'} P(s'|s,a)[R(s,a,s') + \gamma \alpha_{p'}(s')]$$

p = [a, p'], p' subplan

$$V(b) = \max_p b \cdot \alpha_p$$

conditional plans

Sensorless Vacuum Cleaner World

[Modified from RN 3e 17.14] What is the time complexity of d steps of POMDP value iteration for a sensorless environment? Give an upper bound on the number of α-vectors generated in the process.

sensorless

$$\alpha_p(s) = \sum_{s'} P(s'|s,a)[R(s,a,s') + \gamma \cancel{\sum_{e'} P(e'|s')\alpha_{p.e'}(s')}]$$

$$\alpha_p(s) = \sum_{s'} P(s'|s,a)[R(s,a,s') + \gamma \alpha_{p'}(s')]$$

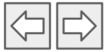
p = [a, p'], p' subplan

$$V(b) = \max_p b \cdot \alpha_p$$

conditional plans

Sensorless Vacuum Cleaner World

depth=1



2

[Modified from RN 3e 17.14] What is the time complexity of d steps of POMDP value iteration for a sensorless environment? Give an upper bound on the number of α-vectors generated in the process.

sensorless

$$\alpha_p(s) = \sum_{s'} P(s'|s,a)[R(s,a,s') + \gamma \cancel{\sum_{e'} P(e'|s')} \alpha_{p.e'}(s')]$$

$$\alpha_p(s) = \sum_{s'} P(s'|s,a)[R(s,a,s') + \gamma \alpha_{p'}(s')]$$

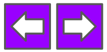
p = [a, p'], p' subplan

$$V(b) = \max_p b \cdot \alpha_p$$

conditional plans

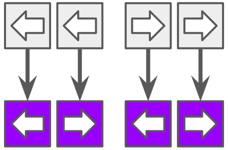
Sensorless Vacuum Cleaner World

depth=1



2

depth=2



2^2

[Modified from RN 3e 17.14] What is the time complexity of d steps of POMDP value iteration for a sensorless environment? Give an upper bound on the number of α-vectors generated in the process.

sensorless

$$\alpha_p(s) = \sum_{s'} P(s'|s,a)[R(s,a,s') + \gamma \cancel{\sum_{e'} P(e'|s')} \alpha_{p.e'}(s')]$$

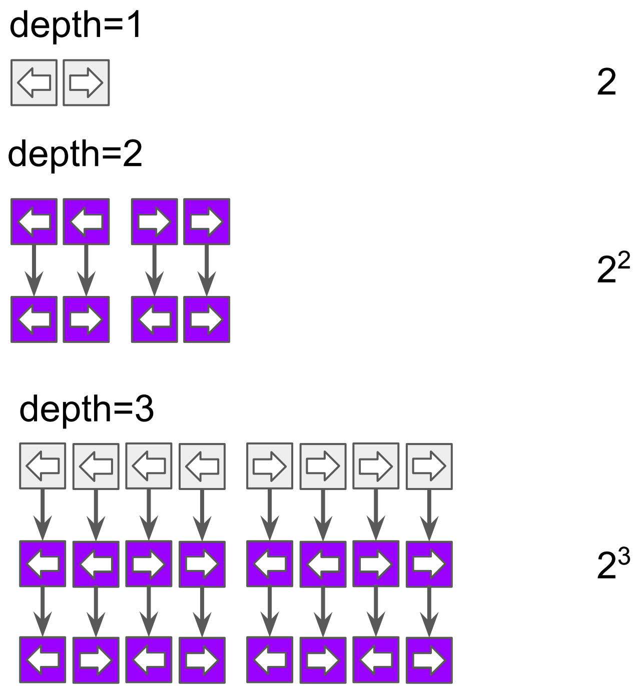
$$\alpha_p(s) = \sum_{s'} P(s'|s,a)[R(s,a,s') + \gamma \alpha_{p'}(s')]$$

p = [a, p'], p' subplan

$$V(b) = \max_p b \cdot \alpha_p$$

conditional plans

Sensorless Vacuum Cleaner World



[Modified from RN 3e 17.14] What is the time complexity of d steps of POMDP value iteration for a sensorless environment? Give an upper bound on the number of α-vectors generated in the process.

sensorless

$$\alpha_p(s) = \sum_{s'} P(s'|s,a)[R(s,a,s') + \gamma \cancel{\sum_{e'} P(e'|s')} \alpha_{p.e'}(s')]$$

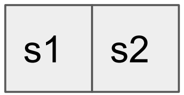
$$\alpha_p(s) = \sum_{s'} P(s'|s,a)[R(s,a,s') + \gamma \alpha_{p'}(s')]$$

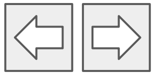
p = [a, p'], p' subplan

$$V(b) = \max_p b \cdot \alpha_p$$

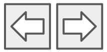
conditional plans

Sensorless Vacuum Cleaner World



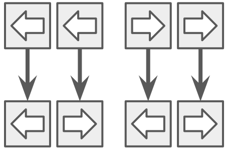


depth=1



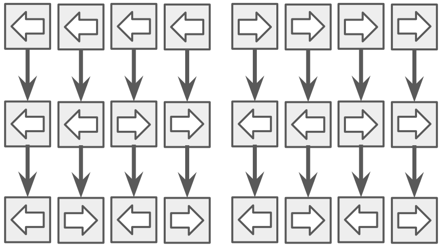
2

depth=2



2^2

depth=3



2^3

depth=d,
|A| actions

$|A|^d$

[Modified from RN 3e 17.14] What is the time complexity of d steps of POMDP value iteration for a sensorless environment? Give an upper bound on the number of α-vectors generated in the process.

sensorless

$$\alpha_p(s) = \sum_{s'} P(s'|s,a)[R(s,a,s') + \gamma \cancel{\sum_{e'} P(e'|s')}\alpha_{p.e'}(s')]$$

$$\alpha_p(s) = \sum_{s'} P(s'|s,a)[R(s,a,s') + \gamma \alpha_{p'}(s')]$$

p = [a, p'], p' subplan

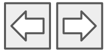
$$V(b) = \max_{p} b \cdot \alpha_p$$

conditional plans

Sensorless Vacuum Cleaner World

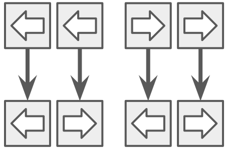
Number of alpha vectors at depth d

depth=1



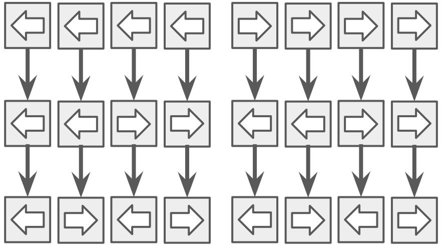
2

depth=2



$+2^2$

depth=3



$+2^3$

depth=d,
|A| actions

$+|A|^d$

[Modified from RN 3e 17.14] What is the time complexity of d steps of POMDP value iteration for a sensorless environment? Give an upper bound on the number of α-vectors generated in the process.

sensorless

$$\alpha_p(s) = \sum_{s'} P(s'|s,a)[R(s,a,s') + \gamma \cancel{\sum_{e'} P(e'|s')}\alpha_{p.e'}(s')]$$

$$\alpha_p(s) = \sum_{s'} P(s'|s,a)[R(s,a,s') + \gamma \alpha_{p'}(s')]$$

p = [a, p'], p' subplan

$$V(b) = \max_p b \cdot \alpha_p$$

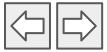
conditional plans

Sensorless Vacuum Cleaner World

Number of alpha vectors at depth d

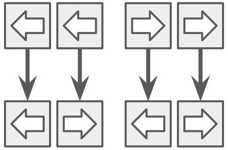
$$\sum_d |A|^d = O(|A|^d)$$

depth=1



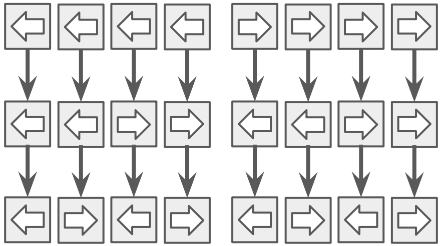
2

depth=2



$+2^2$

depth=3

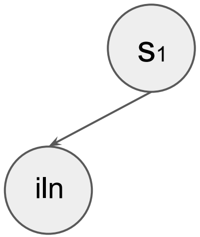


$+2^3$

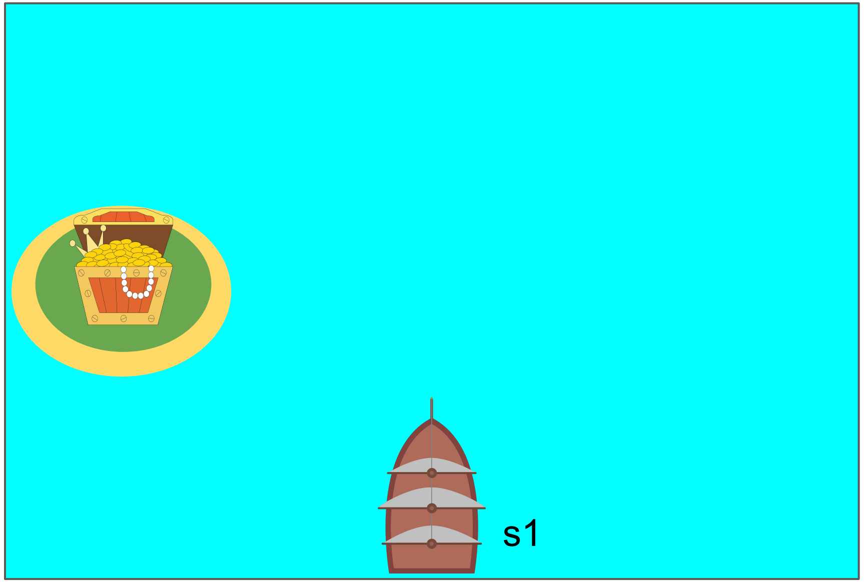
depth=d,
|A| actions

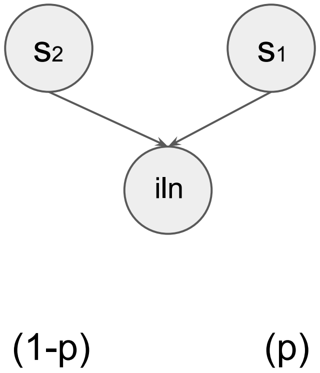
$+|A|^d$

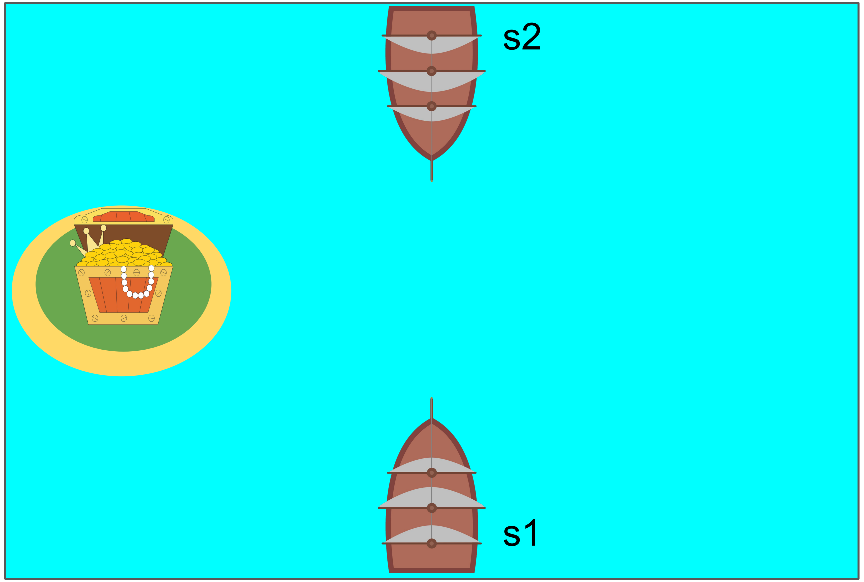
Third

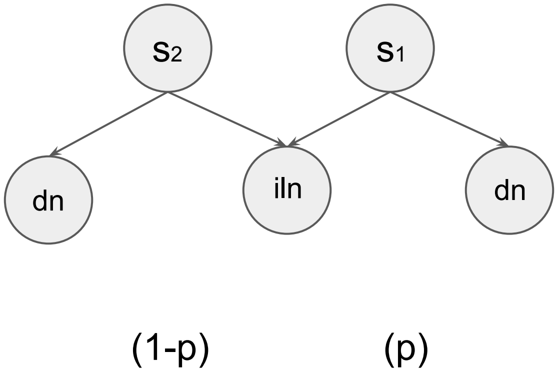


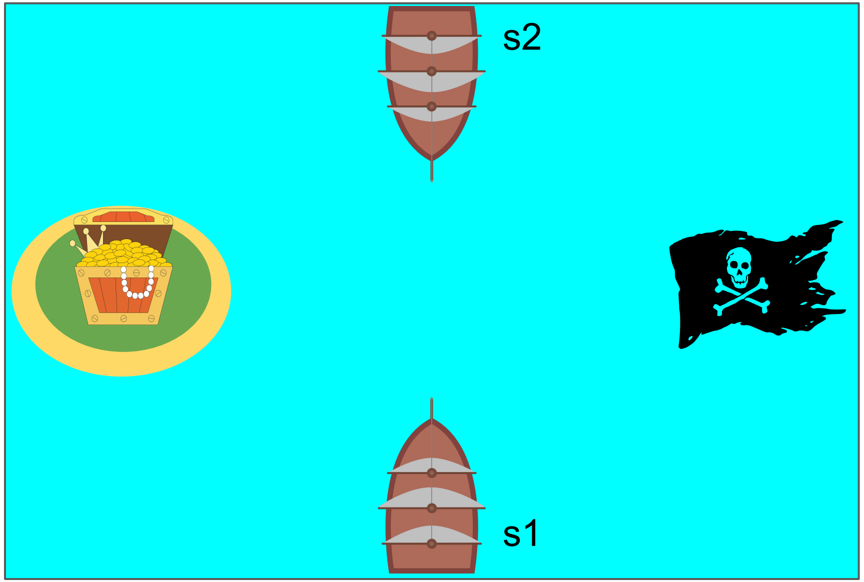
(p)

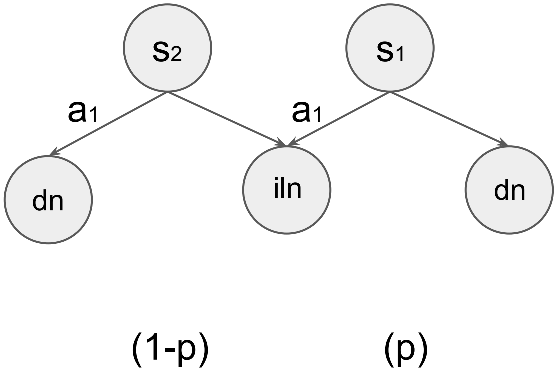


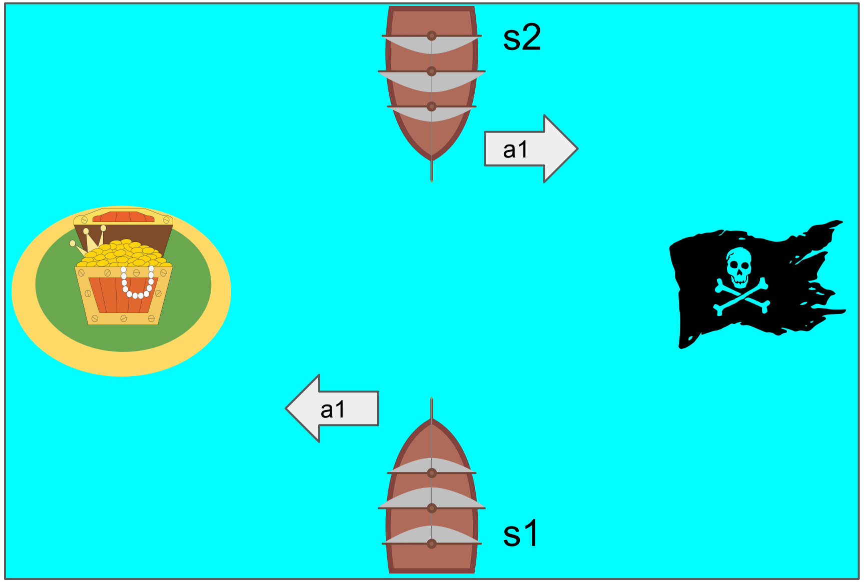

s2
s1
iln
(1-p)
(p)

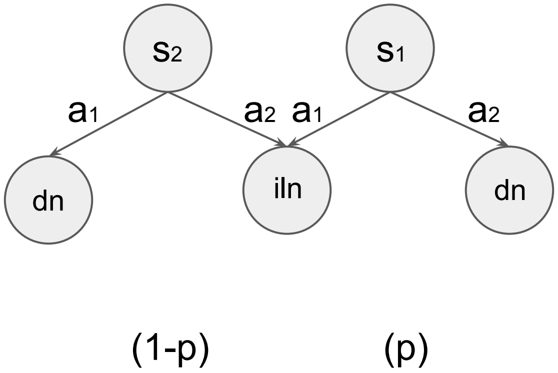

s2
s1


s2
s1
dn
iln
dn
(1-p)
(p)

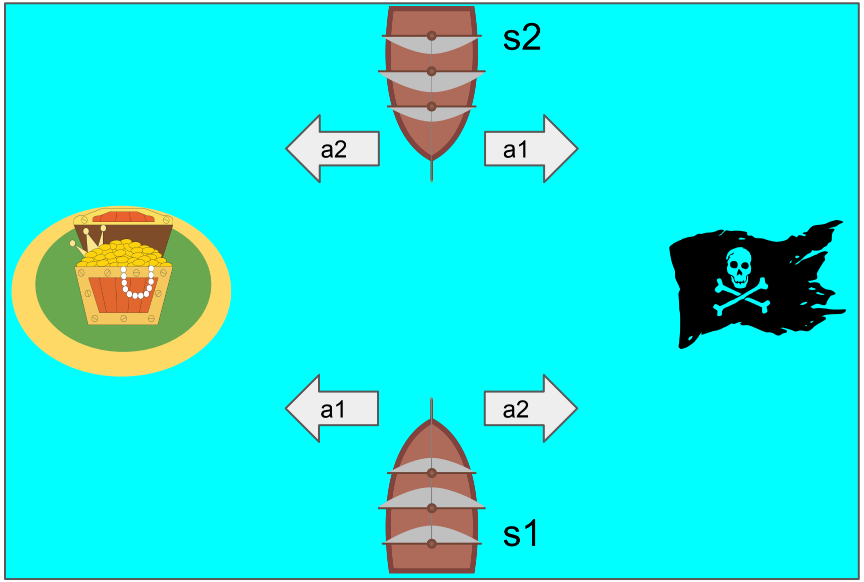

s2
s1


s_2
s_1
a_1
a_1
dn
iln
dn
(1-p)
(p)

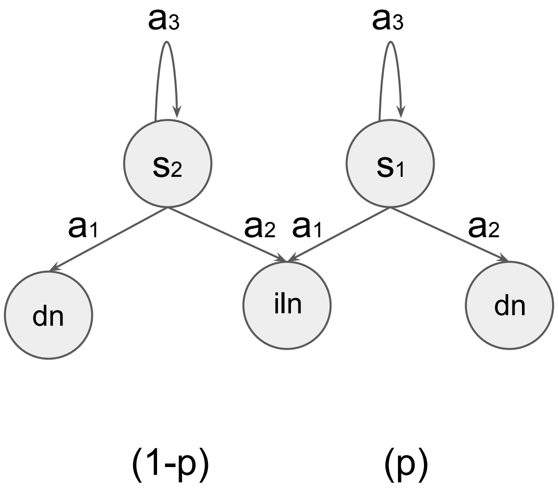

s2
a1
a1
s1



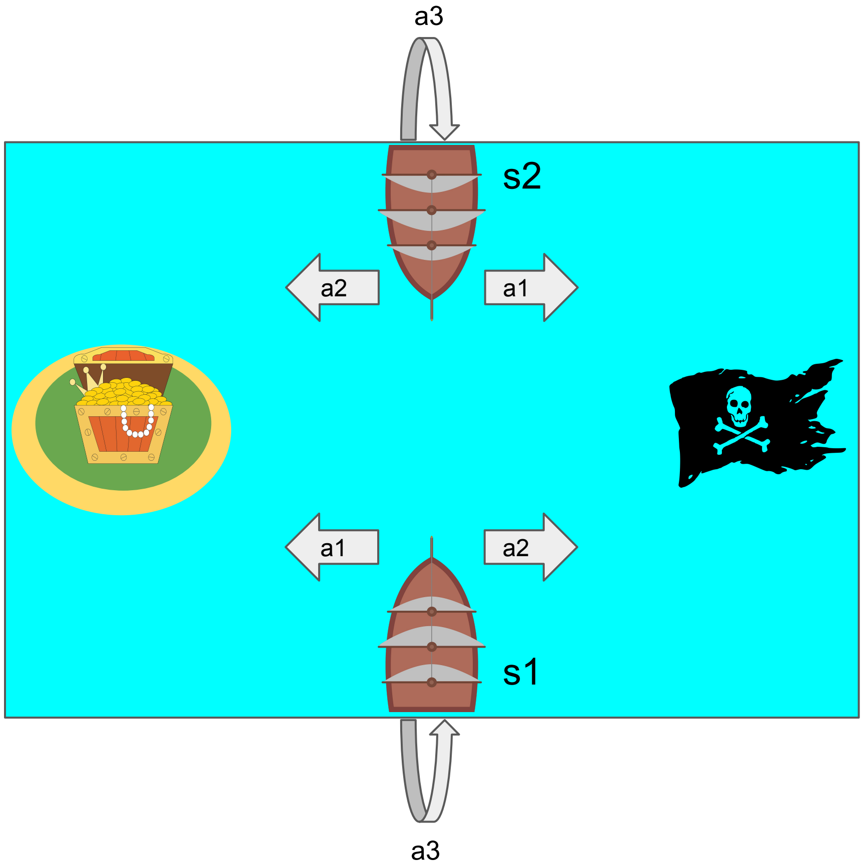
s_2
s_1
a_1
a_2
a_1
a_2
dn
iln
dn
(1-p)
(p)

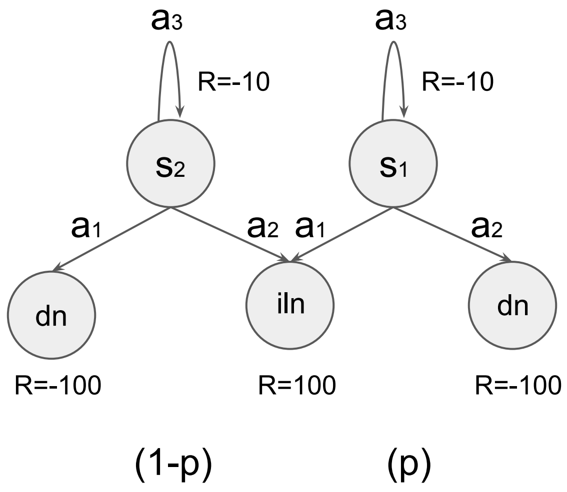


s2
a2
a1
a1
a2
s1

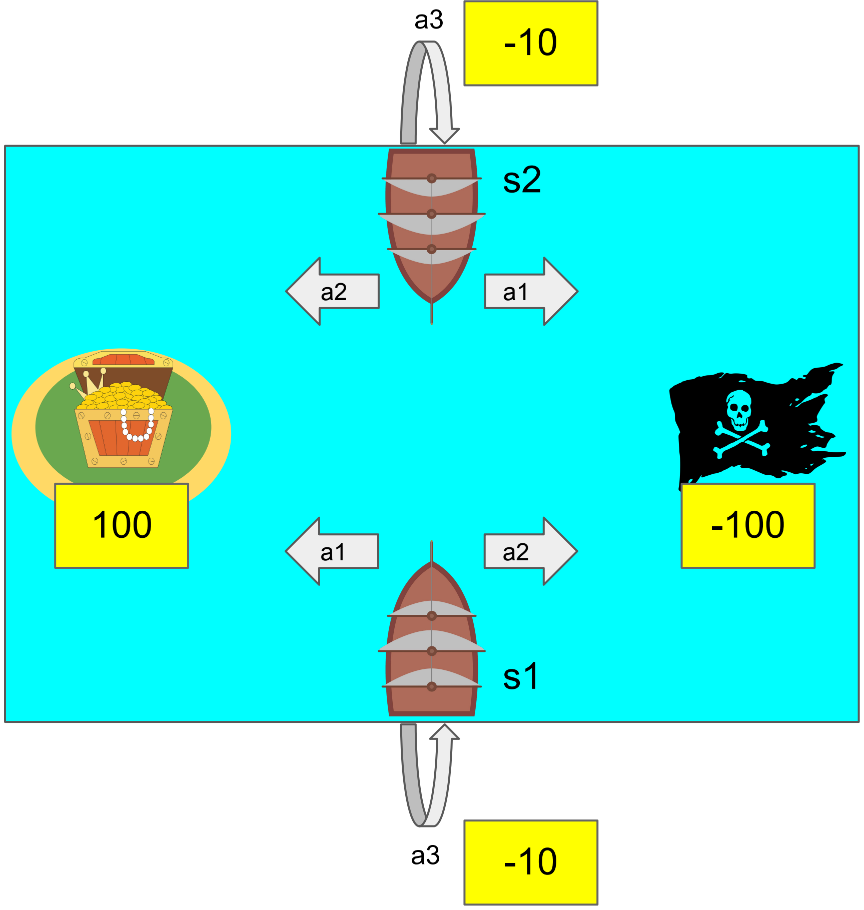


Ask Keeper

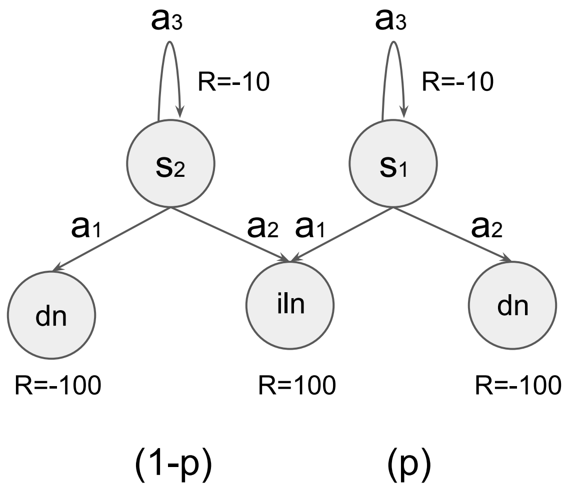




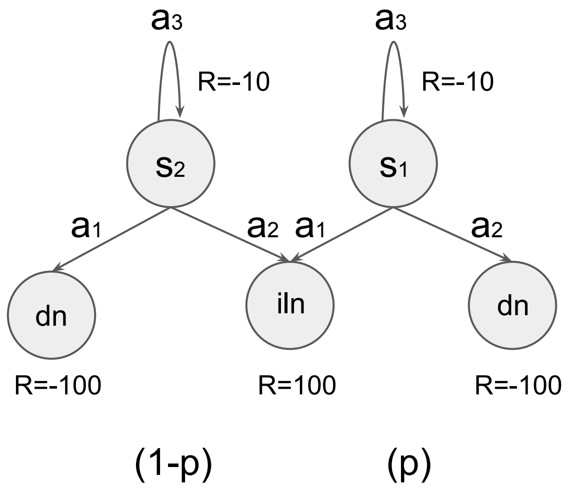
a_3
R=-10
s_2
a_1
a_2
dn
R=-100
iln
R=100
(1-p)
a_3
R=-10
s_1
a_1
a_2
dn
R=-100
(p)



a3
-10
s2
a2
a1
100
-100
a1
a2
s1
a3
-10



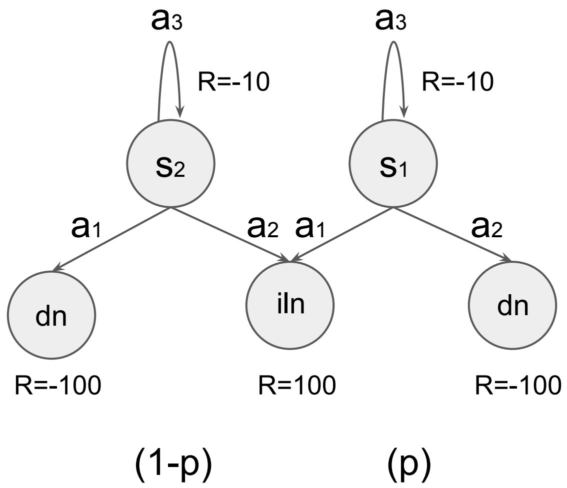
(a) The value of a one-step plan taken in state s is simply the reward of taking the action a in state s: $R(s, a)$. Going left or right are terminal actions while asking the Keeper is non-terminal. Hence, two-step conditional plans can only start with the non-terminal action of asking the Keeper (a_3) followed by an observation and ends with taking another action.



(a) The value of a one-step plan taken in state s is simply the reward of taking the action a in state s: $R(s,a)$. Going left or right are terminal actions while asking the Keeper is non-terminal. Hence, two-step conditional plans can only start with the non-terminal action of asking the Keeper (a_3) followed by an observation and ends with taking another action.

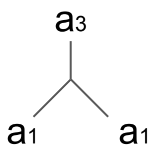
i. How many two-step conditional plans that starts with action a_3 are there?

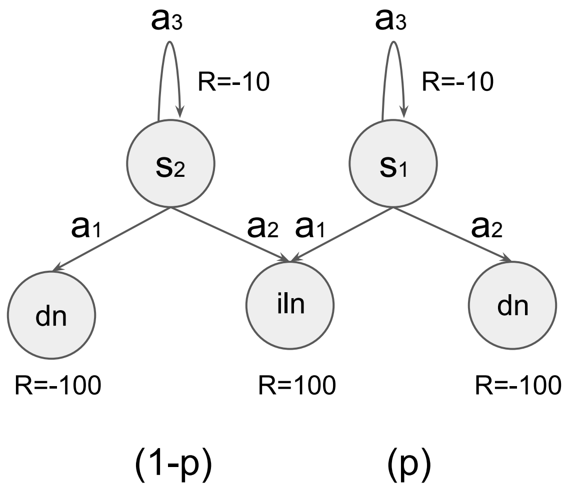
Question



(a) The value of a one-step plan taken in state s is simply the reward of taking the action a in state s: $R(s, a)$. Going left or right are terminal actions while asking the Keeper is non-terminal. Hence, two-step conditional plans can only start with the non-terminal action of asking the Keeper (a_3) followed by an observation and ends with taking another action.

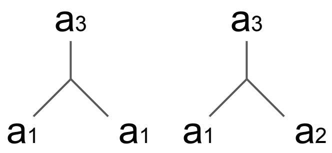
i. How many two-step conditional plans that starts with action a_3 are there?

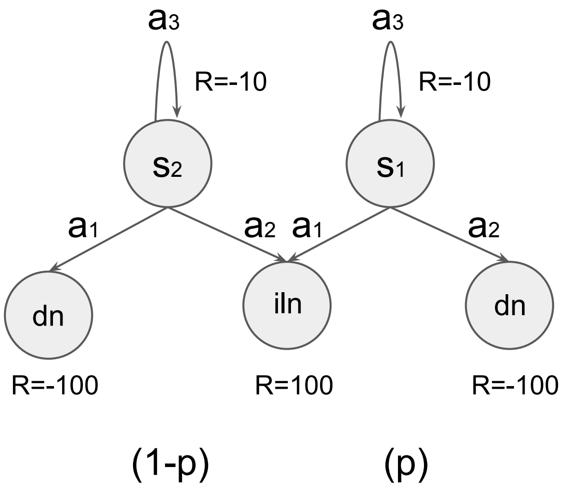




(a) The value of a one-step plan taken in state s is simply the reward of taking the action a in state s: $R(s, a)$. Going left or right are terminal actions while asking the Keeper is non-terminal. Hence, two-step conditional plans can only start with the non-terminal action of asking the Keeper (a_3) followed by an observation and ends with taking another action.

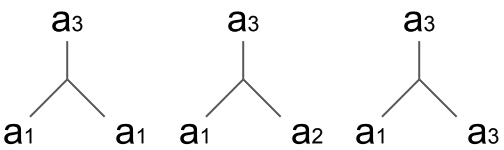
i. How many two-step conditional plans that starts with action a_3 are there?

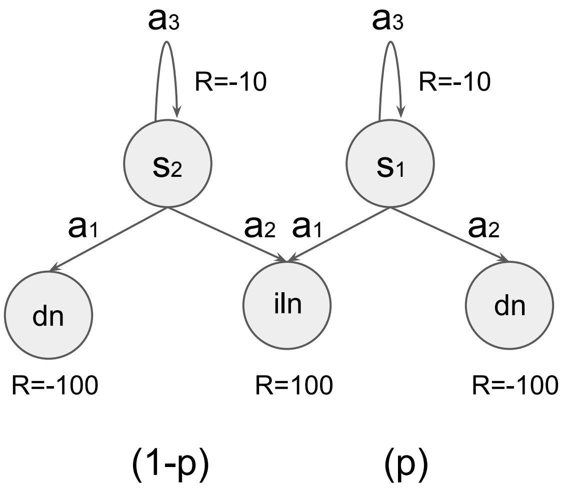




(a) The value of a one-step plan taken in state s is simply the reward of taking the action a in state s: $R(s, a)$. Going left or right are terminal actions while asking the Keeper is non-terminal. Hence, two-step conditional plans can only start with the non-terminal action of asking the Keeper (a_3) followed by an observation and ends with taking another action.

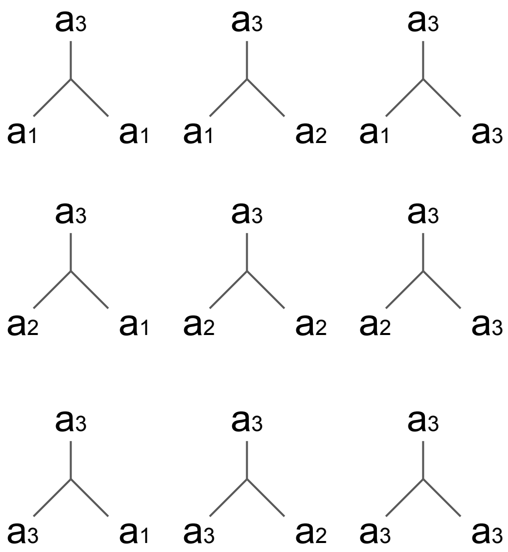
i. How many two-step conditional plans that starts with action a_3 are there?

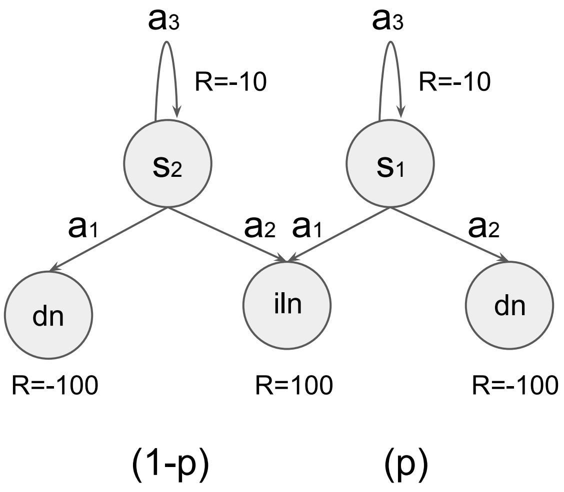




(a) The value of a one-step plan taken in state s is simply the reward of taking the action a in state s: $R(s, a)$. Going left or right are terminal actions while asking the Keeper is non-terminal. Hence, two-step conditional plans can only start with the non-terminal action of asking the Keeper (a_3) followed by an observation and ends with taking another action.

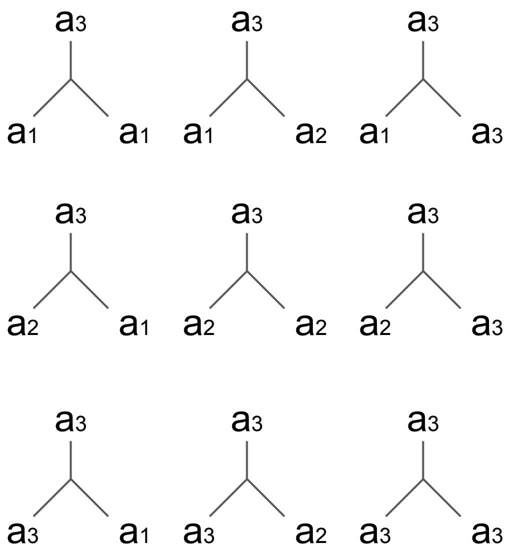
i. How many two-step conditional plans that starts with action a_3 are there?

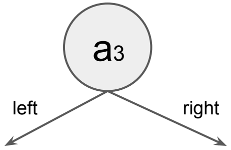


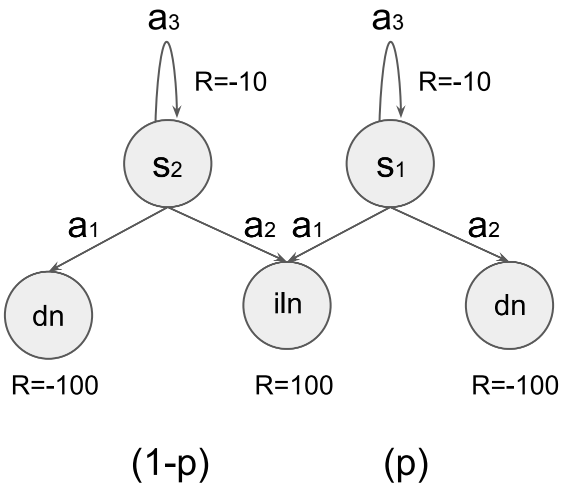


(a) The value of a one-step plan taken in state s is simply the reward of taking the action a in state s: $R(s, a)$. Going left or right are terminal actions while asking the Keeper is non-terminal. Hence, two-step conditional plans can only start with the non-terminal action of asking the Keeper (a_3) followed by an observation and ends with taking another action.

i. How many two-step conditional plans that starts with action a_3 are there?

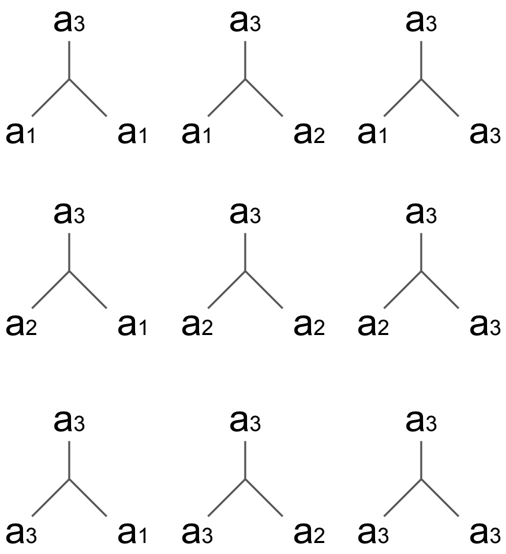


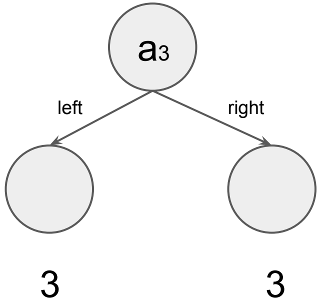


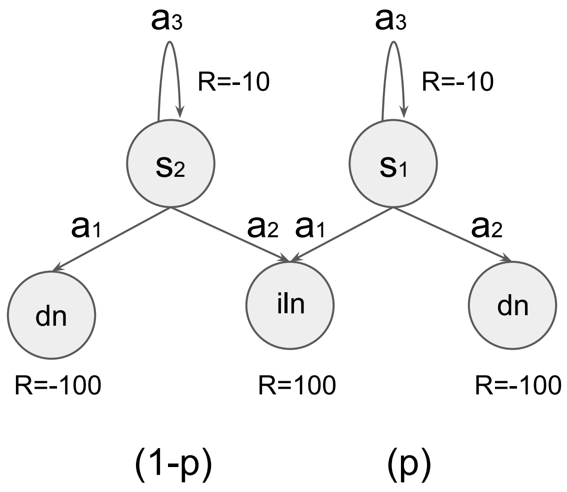


(a) The value of a one-step plan taken in state s is simply the reward of taking the action a in state s: $R(s, a)$. Going left or right are terminal actions while asking the Keeper is non-terminal. Hence, two-step conditional plans can only start with the non-terminal action of asking the Keeper (a_3) followed by an observation and ends with taking another action.

i. How many two-step conditional plans that starts with action a_3 are there?

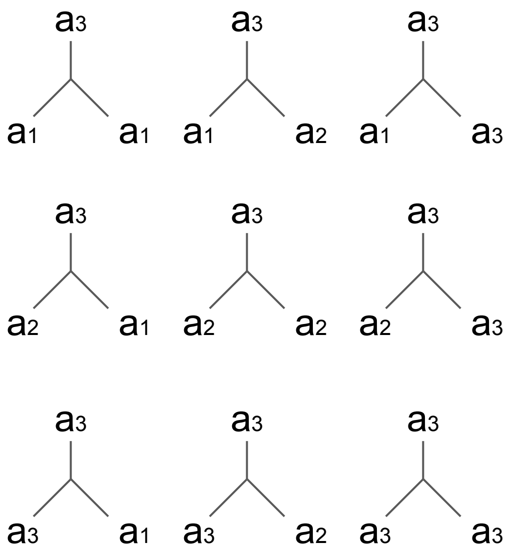


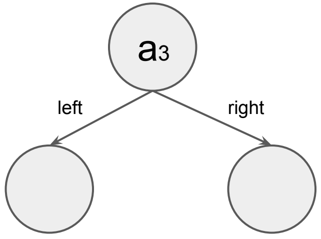




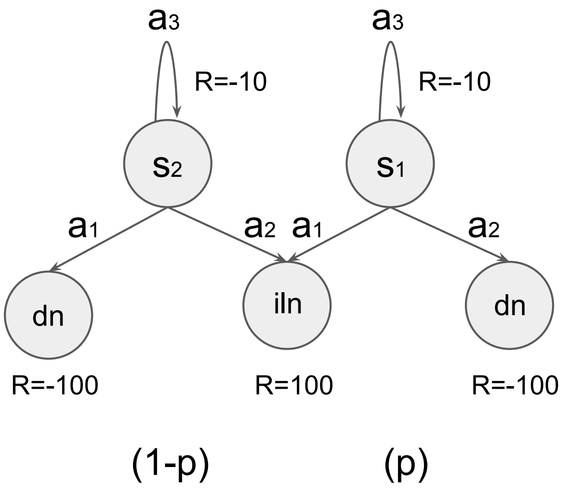
(a) The value of a one-step plan taken in state s is simply the reward of taking the action a in state s: $R(s, a)$. Going left or right are terminal actions while asking the Keeper is non-terminal. Hence, two-step conditional plans can only start with the non-terminal action of asking the Keeper (a_3) followed by an observation and ends with taking another action.

i. How many two-step conditional plans that starts with action a_3 are there?



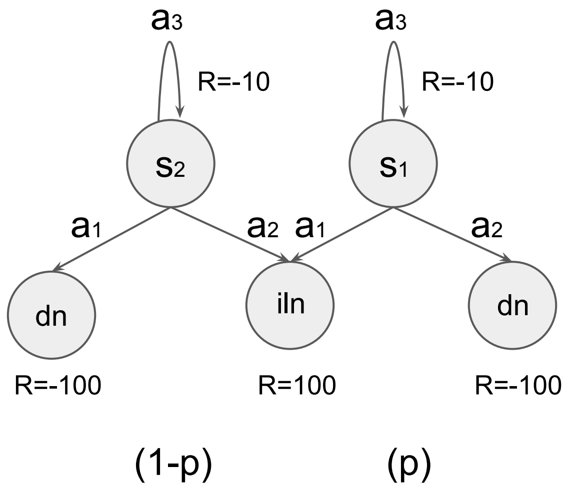


3 x 3 = 9



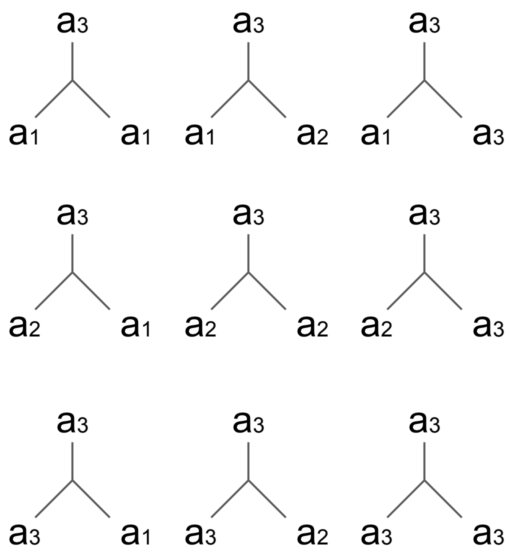
(a) The value of a one-step plan taken in state s is simply the reward of taking the action a in state s: $R(s, a)$. Going left or right are terminal actions while asking the Keeper is non-terminal. Hence, two-step conditional plans can only start with the non-terminal action of asking the Keeper (a_3) followed by an observation and ends with taking another action.

ii. There is only one non-dominated two-step conditional plan: draw (or clearly describe) the non-dominated two step conditional plan.

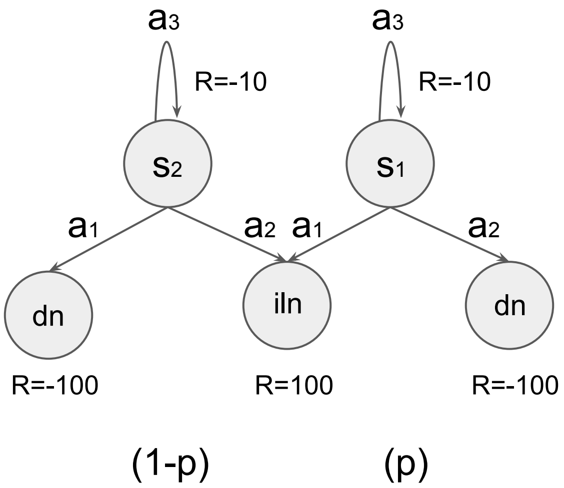


(a) The value of a one-step plan taken in state s is simply the reward of taking the action a in state s: $R(s, a)$. Going left or right are terminal actions while asking the Keeper is non-terminal. Hence, two-step conditional plans can only start with the non-terminal action of asking the Keeper (a_3) followed by an observation and ends with taking another action.

ii. There is only one non-dominated two-step conditional plan: draw (or clearly describe) the non-dominated two step conditional plan.

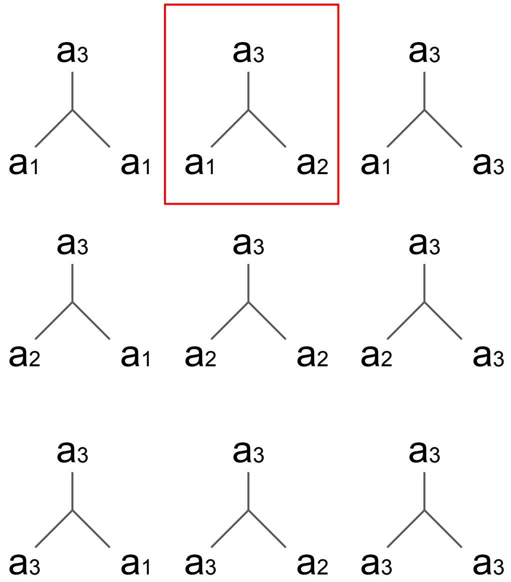


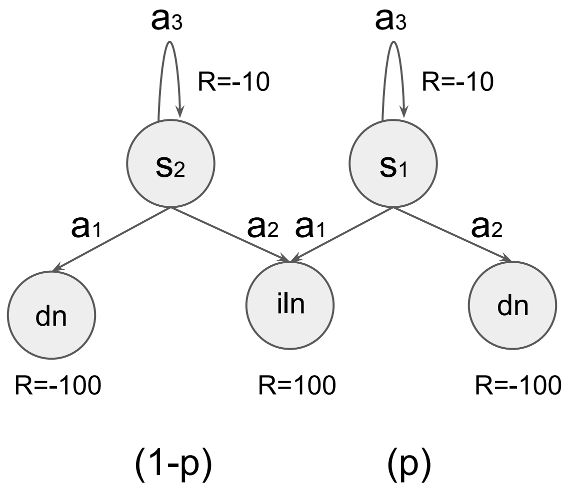
Question



(a) The value of a one-step plan taken in state s is simply the reward of taking the action a in state s: $R(s, a)$. Going left or right are terminal actions while asking the Keeper is non-terminal. Hence, two-step conditional plans can only start with the non-terminal action of asking the Keeper (a_3) followed by an observation and ends with taking another action.

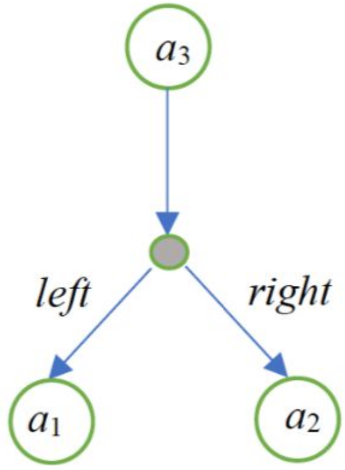
ii. There is only one non-dominated two-step conditional plan: draw (or clearly describe) the non-dominated two step conditional plan.

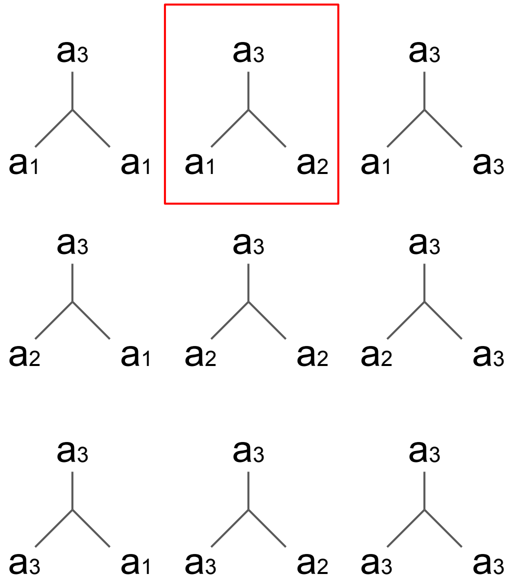


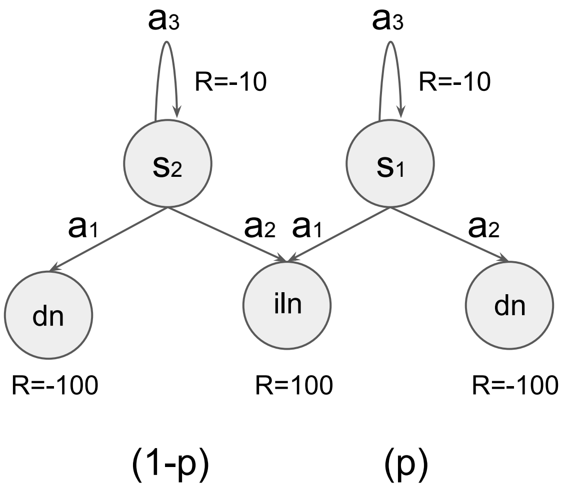


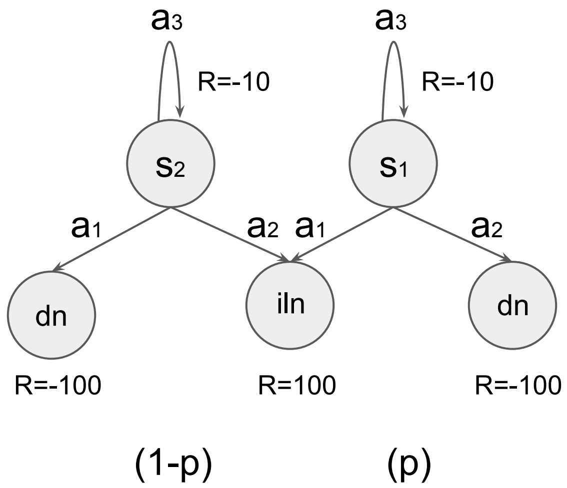
(a) The value of a one-step plan taken in state s is simply the reward of taking the action a in state s: $R(s, a)$. Going left or right are terminal actions while asking the Keeper is non-terminal. Hence, two-step conditional plans can only start with the non-terminal action of asking the Keeper (a_3) followed by an observation and ends with taking another action.

ii. There is only one non-dominated two-step conditional plan: draw (or clearly describe) the non-dominated two step conditional plan.

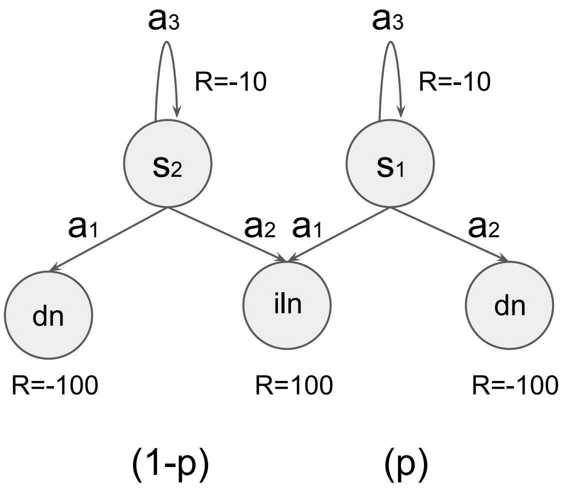





a3
R=-10
s2
a1
a2
dn
R=-100
iln
R=100
a3
R=-10
s1
a1
a2
dn
R=-100
(1-p)
(p)



(b) The one-step plan consisting of asking the Keeper cannot be optimal. Hence there can be at most two non-dominated one-step plans. From part (a) of this question, we know that there is only one non-dominated two-step conditional plan, giving a total of 3 non-dominated one and two step plans.



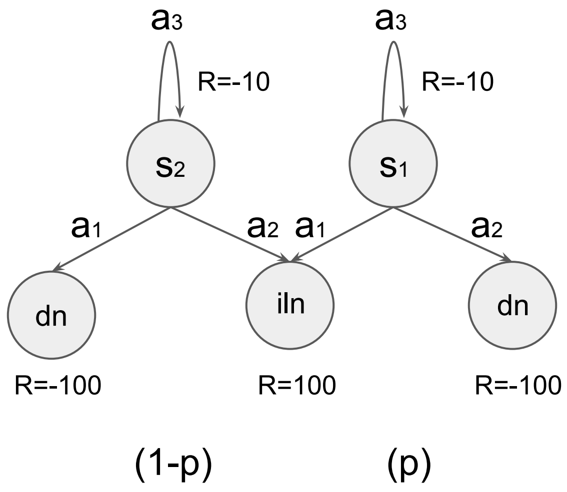
(b) The one-step plan consisting of asking the Keeper cannot be optimal. Hence there can be at most two non-dominated one-step plans. From part (a) of this question, we know that there is only one non-dominated two-step conditional plan, giving a total of 3 non-dominated one and two step plans.

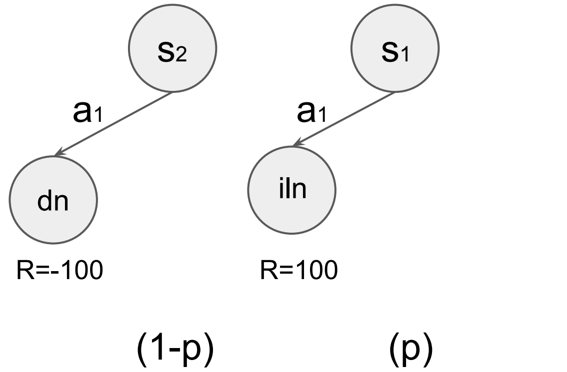
i. Give the three α-vectors corresponding to the three non-dominated plans. Assume that the discount factor is $\gamma = 1$ (not discounted).

Question

(b) The one-step plan consisting of asking the Keeper cannot be optimal. Hence there can be at most two non-dominated one-step plans. From part (a) of this question, we know that there is only one non-dominated two-step conditional plan, giving a total of 3 non-dominated one and two step plans.

i. Give the three α-vectors corresponding to the three non-dominated plans. Assume that the discount factor is $\gamma = 1$ (not discounted).



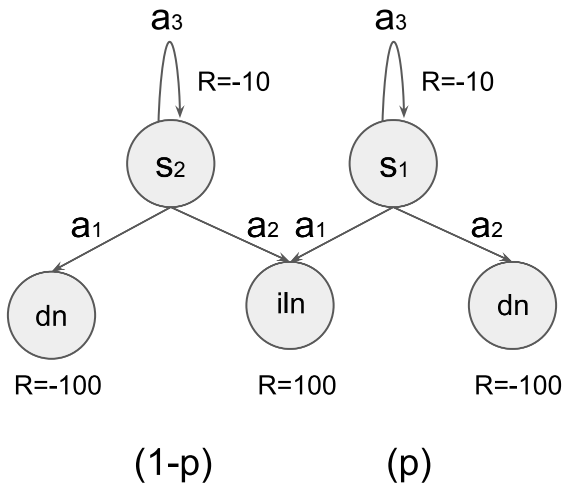


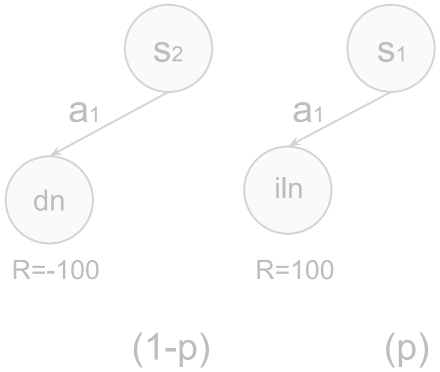
Action left: $\alpha_l(s_1) = R(s_1, a_1) = 100,$

$\alpha_l(s_2) = R(s_2, a_1) = -100$

(b) The one-step plan consisting of asking the Keeper cannot be optimal. Hence there can be at most two non-dominated one-step plans. From part (a) of this question, we know that there is only one non-dominated two-step conditional plan, giving a total of 3 non-dominated one and two step plans.

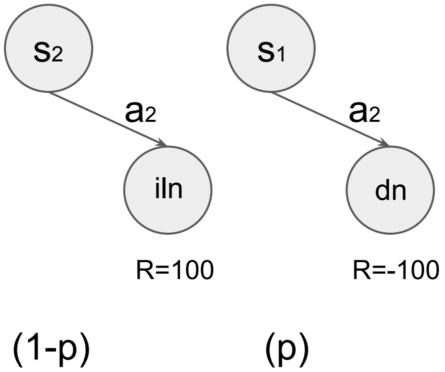
i. Give the three α-vectors corresponding to the three non-dominated plans. Assume that the discount factor is $\gamma = 1$ (not discounted).





Action left: $\alpha_l(s_1) = R(s_1, a_1) = 100,$

$\alpha_l(s_2) = R(s_2, a_1) = -100$

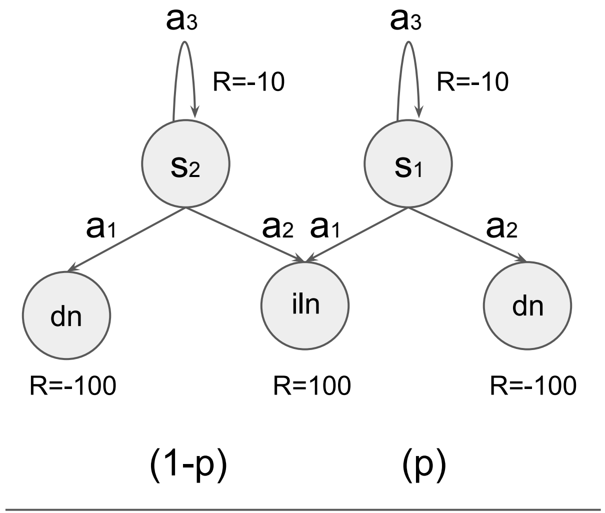


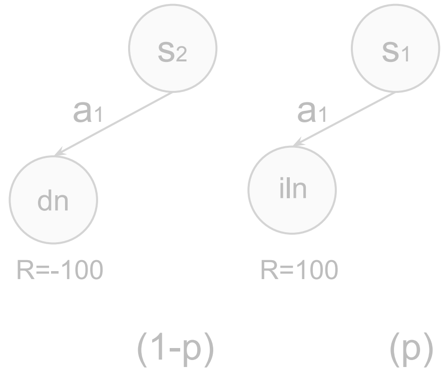
Action right: $\alpha_r(s_1) = R(s_1, a_2) = -100$

$\alpha_l(s_2) = R(s_2, a_2) = 100$

(b) The one-step plan consisting of asking the Keeper cannot be optimal. Hence there can be at most two non-dominated one-step plans. From part (a) of this question, we know that there is only one non-dominated two-step conditional plan, giving a total of 3 non-dominated one and two step plans.

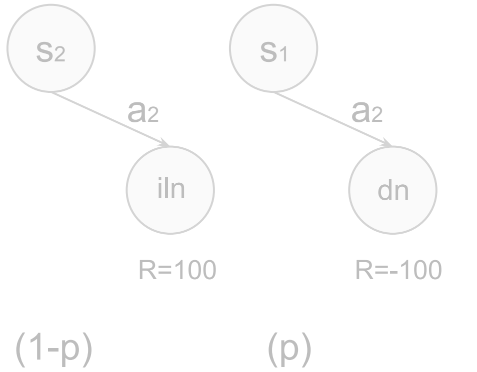
i. Give the three α-vectors corresponding to the three non-dominated plans. Assume that the discount factor is $\gamma = 1$ (not discounted).





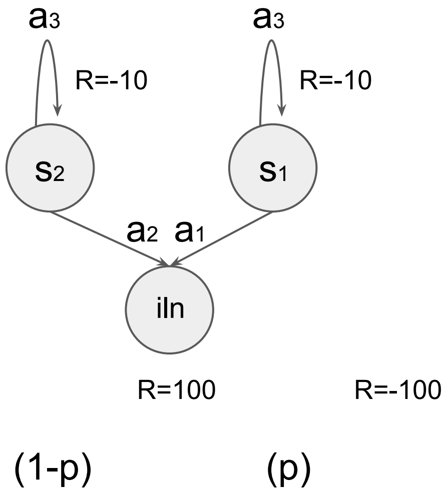
Action left: $\alpha_l(s_1) = R(s_1, a_1) = 100,$

$\alpha_l(s_2) = R(s_2, a_1) = -100$



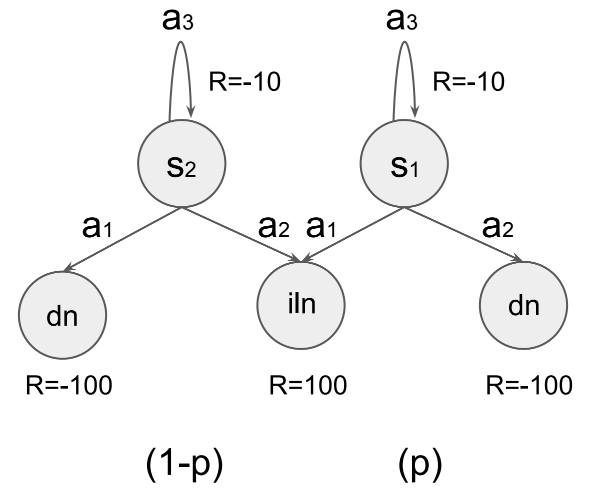
Action right: $\alpha_r(s_1) = R(s_1, a_2) = -100$

$\alpha_l(s_2) = R(s_2, a_2) = 100$



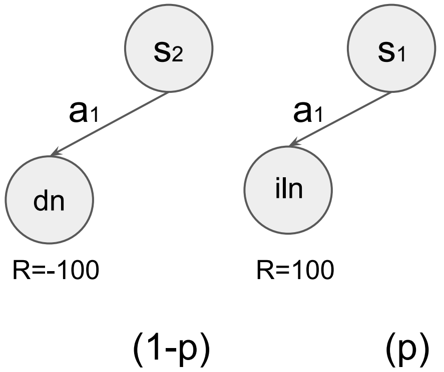
Two-step plan:

$\alpha_p(s_1) = \alpha_p(s_2) = -10 + 100 = 90$



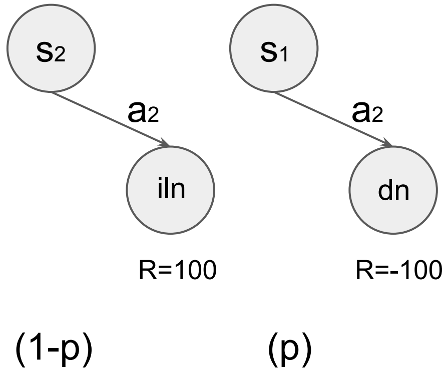
(b) The one-step plan consisting of asking the Keeper cannot be optimal. Hence there can be at most two non-dominated one-step plans. From part (a) of this question, we know that there is only one non-dominated two-step conditional plan, giving a total of 3 non-dominated one and two step plans.

i. Give the three α-vectors corresponding to the three non-dominated plans. Assume that the discount factor is $\gamma = 1$ (not discounted).



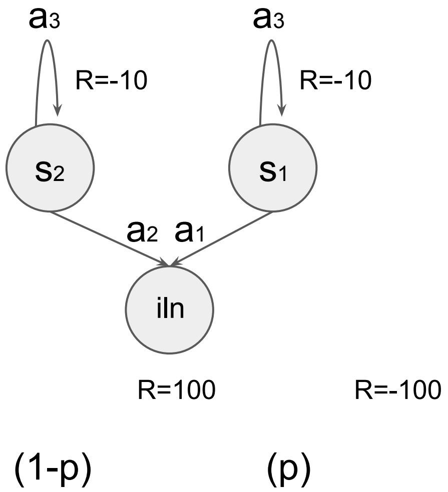
Action left: $\alpha_l(s_1) = R(s_1, a_1) = 100,$

$\alpha_l(s_2) = R(s_2, a_1) = -100$



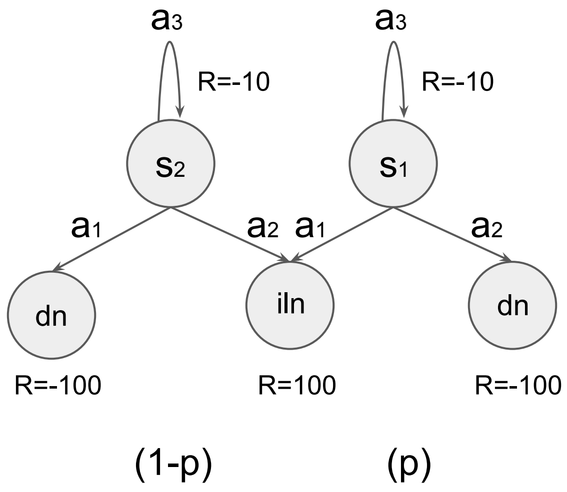
Action right: $\alpha_r(s_1) = R(s_1, a_2) = -100$

$\alpha_l(s_2) = R(s_2, a_2) = 100$

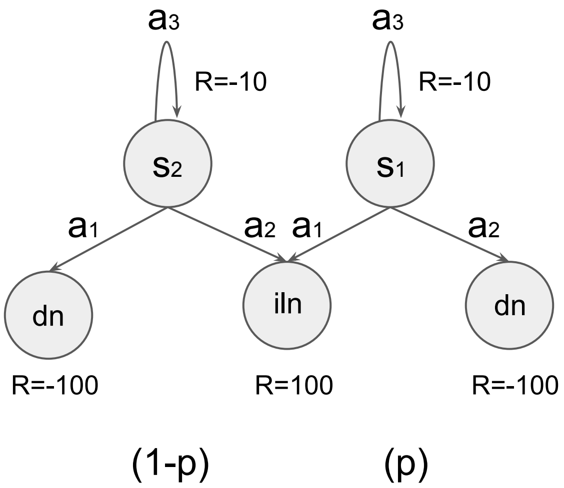


Two-step plan:

$\alpha_p(s_1) = \alpha_p(s_2) = -10 + 100 = 90$



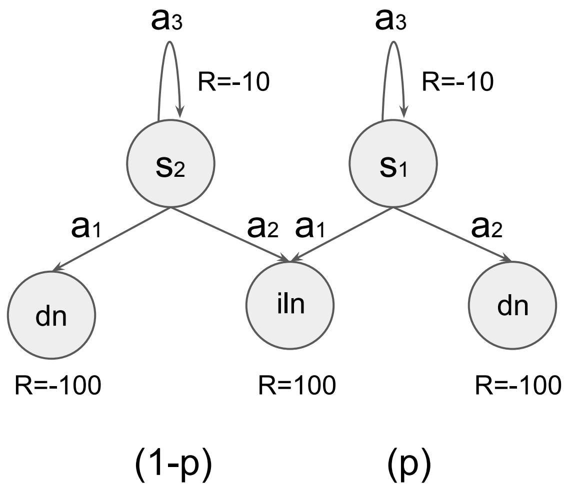
(b) The one-step plan consisting of asking the Keeper cannot be optimal. Hence there can be at most two non-dominated one-step plans. From part (a) of this question, we know that there is only one non-dominated two-step conditional plan, giving a total of 3 non-dominated one and two step plans.



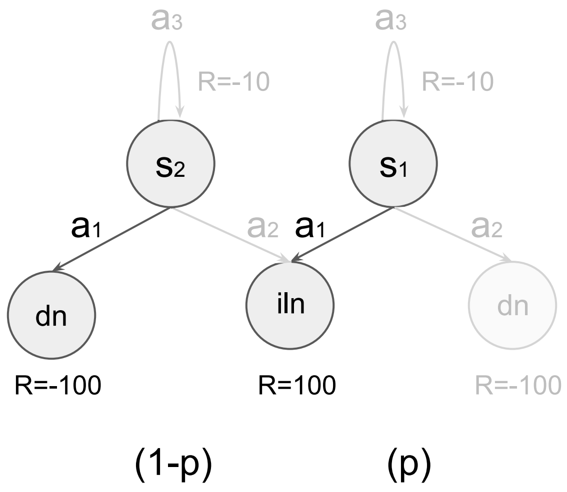
(b) The one-step plan consisting of asking the Keeper cannot be optimal. Hence there can be at most two non-dominated one-step plans. From part (a) of this question, we know that there is only one non-dominated two-step conditional plan, giving a total of 3 non-dominated one and two step plans.

ii. Partition the beliefs into regions where each plan is optimal. Describe the regions.

Question



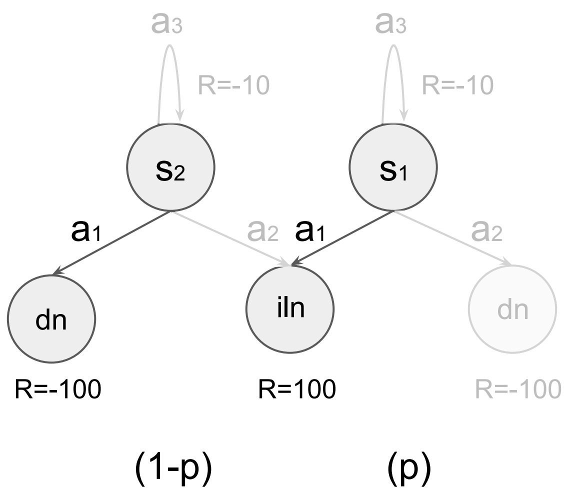
ii. Partition the beliefs into regions where each plan is optimal. Describe the regions.



ii. Partition the beliefs into regions where each plan is optimal. Describe the regions.

Left is optimal:

$$E[\alpha_l] \geq E[\alpha_p]$$

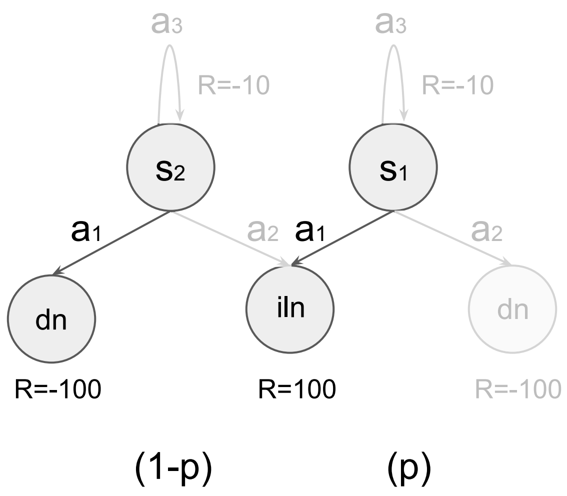


ii. Partition the beliefs into regions where each plan is optimal. Describe the regions.

Left is optimal:

$$E[\alpha_l] \geq E[\alpha_p]$$

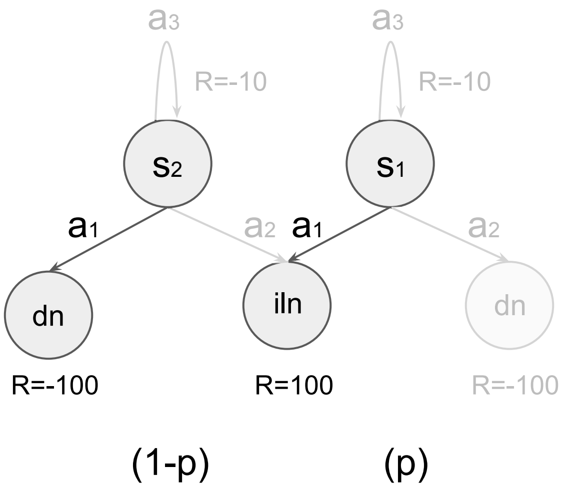
$$p \times \alpha_l(s_1) + (1-p) \times \alpha_l(s_2) \geq 90$$



ii. Partition the beliefs into regions where each plan is optimal. Describe the regions.

Left is optimal:

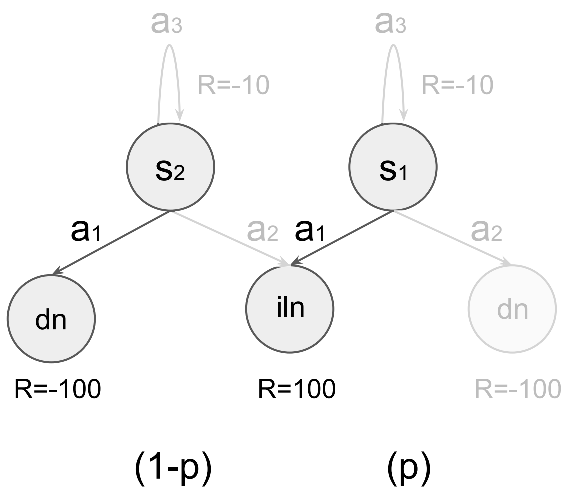
$$E[\alpha_l] \geq E[\alpha_p]$$
$$p \times \alpha_l(s_1) + (1-p) \times \alpha_l(s_2) \geq 90$$
$$p \times 100 + (1-p) \times -100 \geq 90$$



ii. Partition the beliefs into regions where each plan is optimal. Describe the regions.

Left is optimal:

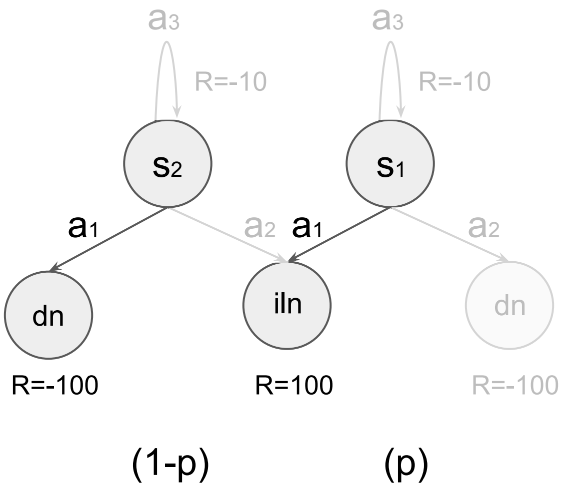
$$E[\alpha_l] \geq E[\alpha_p]$$
$$p \times \alpha_l(s_1) + (1-p) \times \alpha_l(s_2) \geq 90$$
$$p \times 100 + (1-p) \times -100 \geq 90$$
$$100p - 100 + 100p \geq 90$$



ii. Partition the beliefs into regions where each plan is optimal. Describe the regions.

Left is optimal:

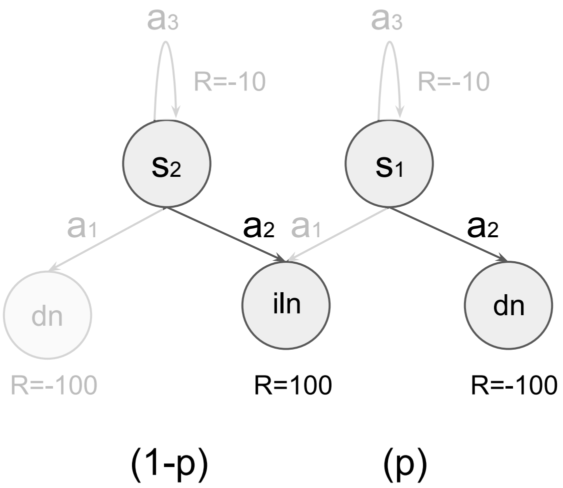
$$
\begin{aligned}
E[\alpha_l] &\geq E[\alpha_p] \\
p \times \alpha_l(s_1) + (1-p) \times \alpha_l(s_2) &\geq 90 \\
p \times 100 + (1-p) \times -100 &\geq 90 \\
100p - 100 + 100p &\geq 90 \\
200p &\geq 190
\end{aligned}
$$



ii. Partition the beliefs into regions where each plan is optimal. Describe the regions.

Left is optimal:

$$
\begin{aligned}
E[\alpha_l] &\geq E[\alpha_p] \\
p \times \alpha_l(s_1) + (1-p) \times \alpha_l(s_2) &\geq 90 \\
p \times 100 + (1-p) \times -100 &\geq 90 \\
100p - 100 + 100p &\geq 90 \\
200p &\geq 190 \\
p &\geq \frac{19}{20}
\end{aligned}
$$

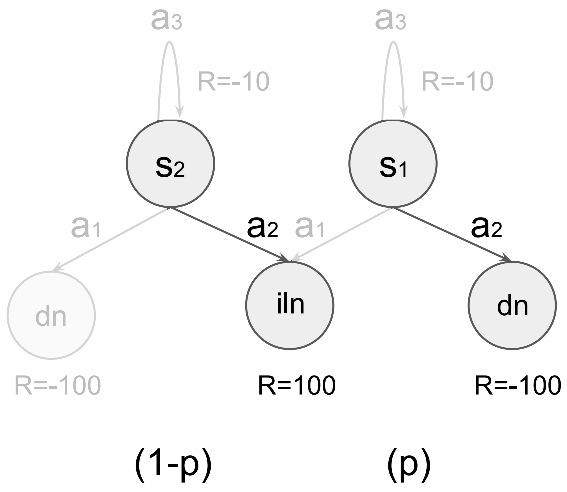


ii. Partition the beliefs into regions where each plan is optimal. Describe the regions.

Left is optimal:

$$\begin{aligned} E[\alpha_l] &\geq E[\alpha_p] \\ p \times \alpha_l(s_1) + (1-p) \times \alpha_l(s_2) &\geq 90 \\ p \times 100 + (1-p) \times -100 &\geq 90 \\ 100p - 100 + 100p &\geq 90 \\ 200p &\geq 190 \\ p &\geq \frac{19}{20} \end{aligned}$$

Right is optimal:



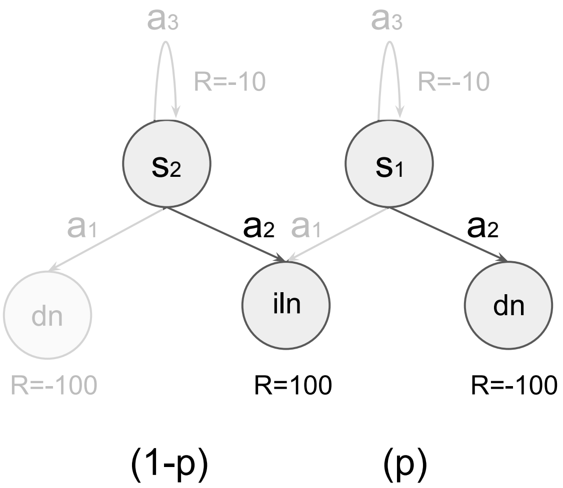
ii. Partition the beliefs into regions where each plan is optimal. Describe the regions.

Left is optimal:

$$\begin{aligned} E[\alpha_l] &\geq E[\alpha_p] \\ p \times \alpha_l(s_1) + (1-p) \times \alpha_l(s_2) &\geq 90 \\ p \times 100 + (1-p) \times -100 &\geq 90 \\ 100p - 100 + 100p &\geq 90 \\ 200p &\geq 190 \\ p &\geq \frac{19}{20} \end{aligned}$$

Right is optimal:

$$E[\alpha_r] \geq E[\alpha_p]$$



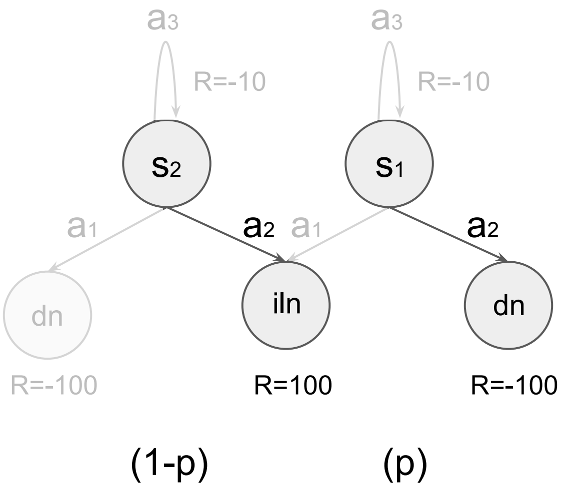
ii. Partition the beliefs into regions where each plan is optimal. Describe the regions.

Left is optimal:

$$
\begin{aligned}
E[\alpha_l] &\geq E[\alpha_p] \\
p \times \alpha_l(s_1) + (1-p) \times \alpha_l(s_2) &\geq 90 \\
p \times 100 + (1-p) \times -100 &\geq 90 \\
100p - 100 + 100p &\geq 90 \\
200p &\geq 190 \\
p &\geq \frac{19}{20}
\end{aligned}
$$

Right is optimal:

$$
\begin{aligned}
E[\alpha_r] &\geq E[\alpha_p] \\
p \times \alpha_r(s_1) + (1-p) \times \alpha_r(s_2) &\geq 90
\end{aligned}
$$



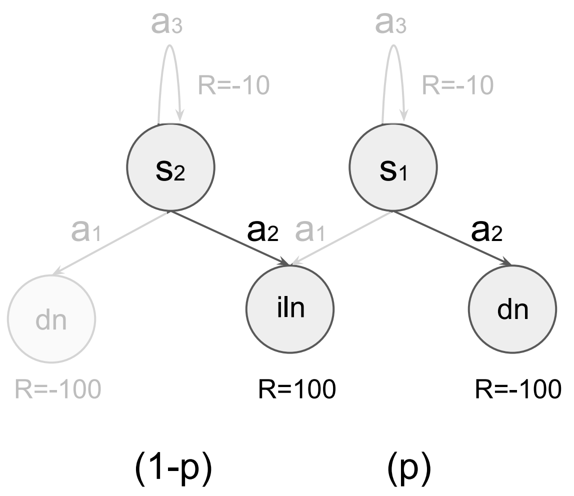
ii. Partition the beliefs into regions where each plan is optimal. Describe the regions.

Left is optimal:

$$
\begin{aligned}
E[\alpha_l] &\geq E[\alpha_p] \\
p \times \alpha_l(s_1) + (1-p) \times \alpha_l(s_2) &\geq 90 \\
p \times 100 + (1-p) \times -100 &\geq 90 \\
100p - 100 + 100p &\geq 90 \\
200p &\geq 190 \\
p &\geq \frac{19}{20}
\end{aligned}
$$

Right is optimal:

$$
\begin{aligned}
E[\alpha_r] &\geq E[\alpha_p] \\
p \times \alpha_r(s_1) + (1-p) \times \alpha_r(s_2) &\geq 90 \\
p \times -100 + (1-p) \times 100 &\geq 90
\end{aligned}
$$



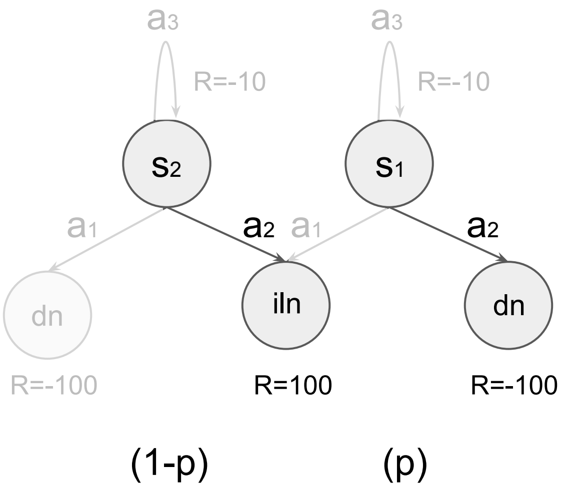
ii. Partition the beliefs into regions where each plan is optimal. Describe the regions.

Left is optimal:

$$
\begin{aligned}
E[\alpha_l] &\geq E[\alpha_p] \\
p \times \alpha_l(s_1) + (1-p) \times \alpha_l(s_2) &\geq 90 \\
p \times 100 + (1-p) \times -100 &\geq 90 \\
100p - 100 + 100p &\geq 90 \\
200p &\geq 190 \\
p &\geq \frac{19}{20}
\end{aligned}
$$

Right is optimal:

$$
\begin{aligned}
E[\alpha_r] &\geq E[\alpha_p] \\
p \times \alpha_r(s_1) + (1-p) \times \alpha_r(s_2) &\geq 90 \\
p \times -100 + (1-p) \times 100 &\geq 90 \\
-100p + 100 - 100p &\geq 90
\end{aligned}
$$



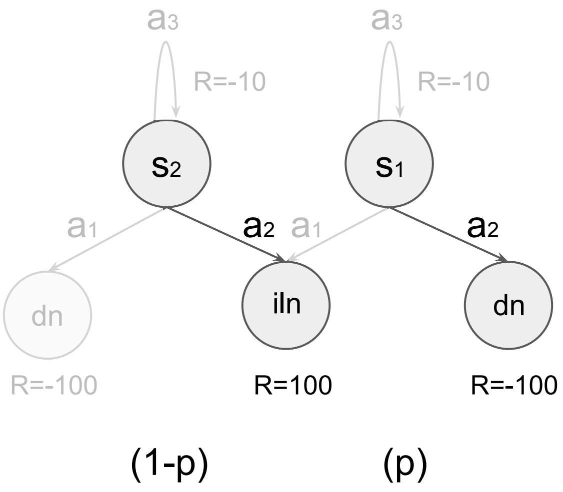
ii. Partition the beliefs into regions where each plan is optimal. Describe the regions.

Left is optimal:

$$\begin{aligned} E[\alpha_l] &\geq E[\alpha_p] \\ p \times \alpha_l(s_1) + (1-p) \times \alpha_l(s_2) &\geq 90 \\ p \times 100 + (1-p) \times -100 &\geq 90 \\ 100p - 100 + 100p &\geq 90 \\ 200p &\geq 190 \\ p &\geq \frac{19}{20} \end{aligned}$$

Right is optimal:

$$\begin{aligned} E[\alpha_r] &\geq E[\alpha_p] \\ p \times \alpha_r(s_1) + (1-p) \times \alpha_r(s_2) &\geq 90 \\ p \times -100 + (1-p) \times 100 &\geq 90 \\ -100p + 100 - 100p &\geq 90 \\ -200p &\geq -10 \end{aligned}$$



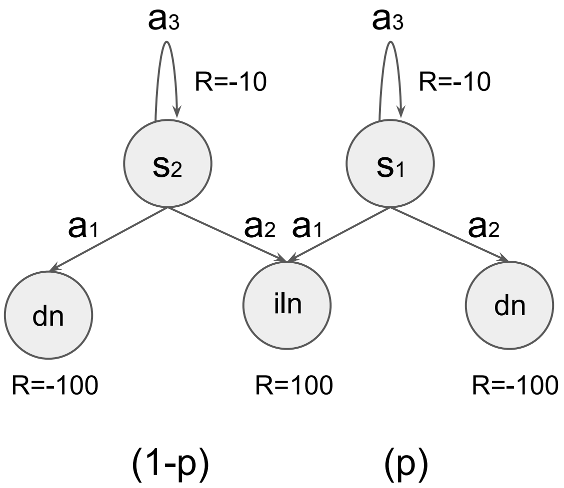
ii. Partition the beliefs into regions where each plan is optimal. Describe the regions.

Left is optimal:

$$\begin{aligned} E[\alpha_l] &\geq E[\alpha_p] \\ p \times \alpha_l(s_1) + (1-p) \times \alpha_l(s_2) &\geq 90 \\ p \times 100 + (1-p) \times -100 &\geq 90 \\ 100p - 100 + 100p &\geq 90 \\ 200p &\geq 190 \\ p &\geq \frac{19}{20} \end{aligned}$$

Right is optimal:

$$\begin{aligned} E[\alpha_r] &\geq E[\alpha_p] \\ p \times \alpha_r(s_1) + (1-p) \times \alpha_r(s_2) &\geq 90 \\ p \times -100 + (1-p) \times 100 &\geq 90 \\ -100p + 100 - 100p &\geq 90 \\ -200p &\geq -10 \\ p &\leq \frac{1}{20} \end{aligned}$$



ii. Partition the beliefs into regions where each plan is optimal. Describe the regions.

Left is optimal:

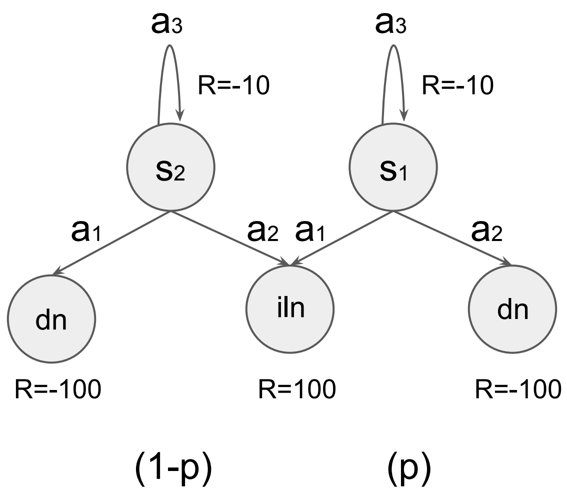
$$\begin{aligned} E[\alpha_l] &\geq E[\alpha_p] \\ p \times \alpha_l(s_1) + (1-p) \times \alpha_l(s_2) &\geq 90 \\ p \times 100 + (1-p) \times -100 &\geq 90 \\ 100p - 100 + 100p &\geq 90 \\ 200p &\geq 190 \\ p &\geq \frac{19}{20} \end{aligned}$$

Two-step is optimal:

$$\frac{1}{20} \leq p \leq \frac{19}{20}$$
$$0.05 \leq p \leq 0.95$$

Right is optimal:

$$\begin{aligned} E[\alpha_r] &\geq E[\alpha_p] \\ p \times \alpha_r(s_1) + (1-p) \times \alpha_r(s_2) &\geq 90 \\ p \times -100 + (1-p) \times 100 &\geq 90 \\ -100p + 100 - 100p &\geq 90 \\ -200p &\geq -10 \\ p &\leq \frac{1}{20} \end{aligned}$$



ii. Partition the beliefs into regions where each plan is optimal. Describe the regions.

Left is optimal:

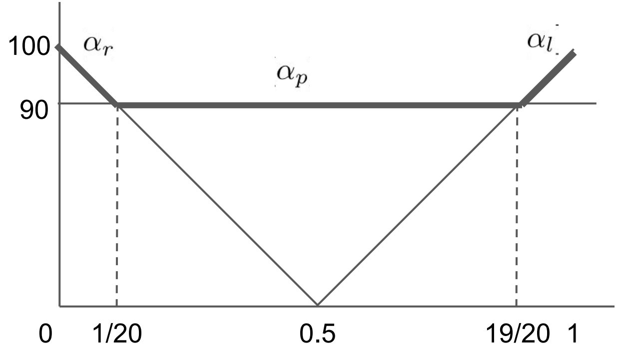
$$\begin{aligned} E[\alpha_l] &\geq E[\alpha_p] \\ p \times \alpha_l(s_1) + (1-p) \times \alpha_l(s_2) &\geq 90 \\ p \times 100 + (1-p) \times -100 &\geq 90 \\ 100p - 100 + 100p &\geq 90 \\ 200p &\geq 190 \\ p &\geq \frac{19}{20} \end{aligned}$$

Right is optimal:

$$\begin{aligned} E[\alpha_r] &\geq E[\alpha_p] \\ p \times \alpha_r(s_1) + (1-p) \times \alpha_r(s_2) &\geq 90 \\ p \times -100 + (1-p) \times 100 &\geq 90 \\ -100p + 100 - 100p &\geq 90 \\ -200p &\geq -10 \\ p &\leq \frac{1}{20} \end{aligned}$$

Two-step is optimal:

$$\frac{1}{20} \leq p \leq \frac{19}{20}$$
$$0.05 \leq p \leq 0.95$$



Question?

<EOF>

Credits

Images are taken from pixabay.com