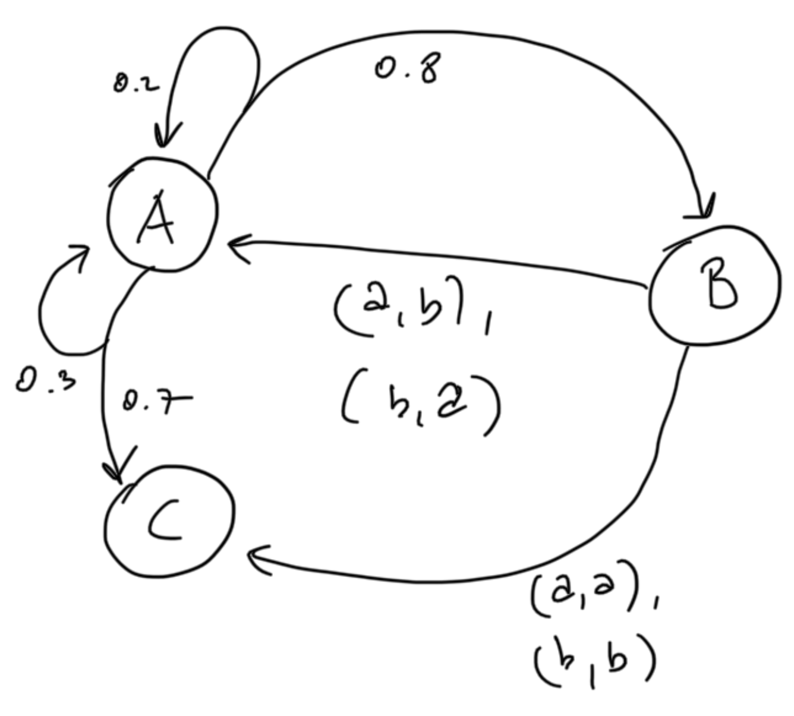


$\gamma = 0.8$
$\alpha = 0.5$

ZERO-SUM GAME
↓
NASH-Q REDUCES TO MINIMAX-Q

A

	a	b
a	0	-1
b	-1	0

B

	a	b
a	-2	0
b	0	-2

C

	a	b
a	1	-1
b	-2	0

ADOPT THE GREEDY POLICY
(WITHOUT EXPLORATION)

A	0	1	2	3	4
a a	0	0	0	0	0
a b	0	0	0	0	0
b a	0	0	0	0	0
b b	0	0	0	0	0

[redacted]

B	0	1	2	3	4
a a	0	0	-1	-1	-1
a b	0	0	0	0	0
b a	0	0	0	0	0
b b	0	0	0	0	-1

[redacted]

C	0	1	2	3	4
a a	0	0	0	0	0
a b	0	0	0	0	0
b a	0	0	0	0	0
b b	0	0	0	0	0

STARTING STATE A

t=1
ACTIONS PLAYED AT A : (a, a) ⟶ B

$$Q(A,a,a) \leftarrow (1-\alpha)\,Q(A,a,a)+\alpha\left(r(A,a,a)+\gamma\, NASHQ(B)\right.$$

0 0.5 0 0.5 0 0.8 0

t=2
ACTIONS PLAYED AT B : (a, a) ⟶ A

$$Q(B,a,a) \leftarrow (1-\alpha)\,Q(B,a,a)+\alpha\left(r(B,a,a)+\gamma\, NASHQ(A)\right.$$

-1 0.5 0 0.5 -2 0.8 0

t=3
ACTION PLAYED AT A : (a, a) ⟶ B

$$Q(A,a,a) \leftarrow (1-\alpha)\,Q(A,a,a)+\alpha\left(r(A,a,a)+\gamma\, NASHQ(B)\right)$$

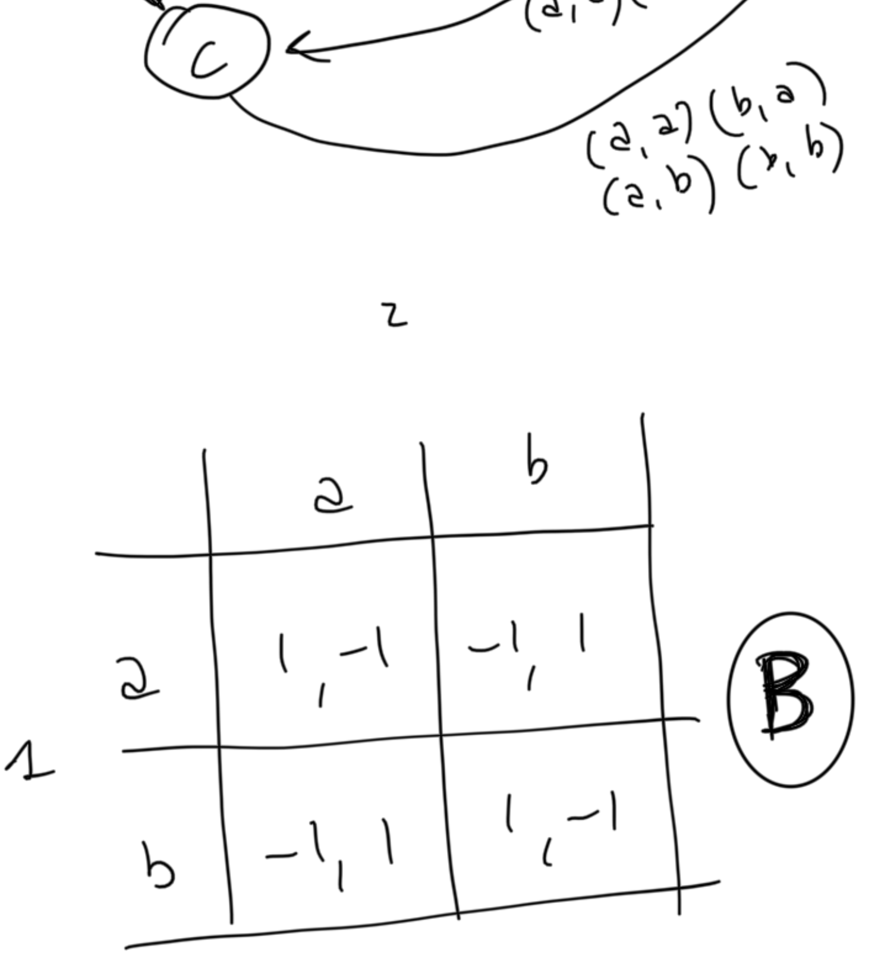
0 0.5 0 0.5 0 0.8 0

↑
PLAYING b IS A DOMINANT STRATEGY FOR PLAYER 1

t=4
ACTION PLAYED AT B : (b, b) ⟶ A

$$Q(B,b,b) \leftarrow (1-\alpha)\,Q(B,b,b)+\alpha\left(r(B,b,b)+\gamma\, NASHQ(A)\right)$$

-1 0.5 0 0.5 -2 0.8 0



		2	
		a	b
1	a	2, 1	0, 0
	b	0, 0	1, 2

Ⓐ

		2	
		a	b
1	a	1, -1	-1, 1
	b	-1, 1	1, -1

Ⓑ

		2	
		a	b
1	a	3, 3	0, 4
	b	4, 0	2, 2

Ⓒ

ADOPT THE GREEDY STRATEGY
(WITHOUT EXPLORATION)

$\alpha = 0.5$ $\gamma = 0.5$

A	0	1	2	3	4
aa	0, 0	1, 0.5	1, 0.5	1, 0.5	1, 0.5
ab	0, 0	0, 0	0, 0	0, 0	0, 0
bc	0, 0	0, 0	0, 0	0, 0	0, 0
bb	0, 0	0, 0	0, 0	0, 0	0, 0

[redacted]

B	0	1	2	3	4
a a	0, 0	0, 0	0, 0	0.875, -0.375	0.875, -0.375
a b	0, 0	0, 0	0, 0	0, 0	0, 0
b a	0, 0	0, 0	0, 0	0, 0	0, 0
b b	0, 0	0, 0	0, 0	0, 0	0, 0

[redacted]

C	0	1	2	3	4
aa	0, 0	0, 0	1.5, 1.5	1.5, 1.5	2.25, 2.25
ab	0, 0	0, 0	0, 0	0, 0	0, 0
ba	0, 0	0, 0	0, 0	0, 0	0, 0
bb	0, 0	0, 0	0, 0	0, 0	0, 0

STARTING STATE A

t=1
ACTIONS PLAYED AT A: (a, a) ⟶ C

$$Q^i(A,a,a) \leftarrow (1-\alpha)\,Q^i(A,a,a)+\alpha\left(r^i(A,a,a)+\gamma\, NASH\text{-}Q^i(C)\right)$$

$i=1$ 1 0.5 0 0.5 2 0.5 0
$i=2$ 0.5 0.5 0 0.5 1 0.5 0

t=2
ACTIONS PLAYED AT C : (a, a) ⟶ B

$$Q^i(C,a,a) \leftarrow (1-\alpha)\,Q^i(C,a,a)+\alpha\left(r^i(C,a,a)+\gamma\, NASHQ^i(B)\right)$$

$i=1$ 1.5 0.5 0 0.5 3 0.5 0
$i=2$ 1.5 0.5 0 0.5 3 0.5 0

t=3
ACTIONS PLAYED AT B: (a, a) ⟶ C

$$Q^i(B,a,a) \leftarrow (1-\alpha)\,Q^i(B,a,a)+\alpha\left(r^i(B,a,a)+\gamma\, NASHQ^i(C)\right)$$

$i=1$ 0.875 0.5 0 0.5 1 0.5 1.5
$i=2$ -0.375 0.5 0 0.5 -1 0.5 1.5

t=4
ACTIONS PLAYED AT C : (a, a) ⟶ B

$$Q^i(C,a,a) \leftarrow (1-\alpha)\,Q^i(C,a,a)+\alpha\left(r^i(C,a,a)+\gamma\, NASH\,Q^i(B)\right)$$

$i=1$ 2.25 0.5 1.5 0.5 3 0.5 0
$i=2$ 2.25 0.5 1.5 0.5 3 0.5 0