# Discount factor influence to policy estimation analysis J. Kostlivá, Z. Straka, P. Švarný 

We have:

- an unknown grid world of unknown size and structure
- robot/agents moves in unknown directions with unknown parameters
- a few episodes the robot tried


## Discount factor influence to policy estimation analysis J. Kostlivá, Z. Straka, P. Švarný

We have:

- an unknown grid world of unknown size and structure
- robot/agents moves in unknown directions with unknown parameters
- a few episodes the robot tried

Today:

- We will compute the optimal policy
- Use different $\gamma$ settings
- Study the boundary values for $\gamma$


## Example I

## Example I

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

## Example I

| Episode 1 Episode 2 Episode 3 Episode 4 <br> $(B, \rightarrow, C,-1)$ $(B, \leftarrow, A,-1)$ $(C, \rightarrow, D,-1)$ $(C, \leftarrow, B,-1)$ <br> $(C, \rightarrow, D,-1)$ $(A, \rightarrow$, exit, 10) $(D, \rightarrow$, exit, 6) $(B, \rightarrow, C,-1)$ <br> $(D, \leftarrow$, exit, 6)   $(C, \leftarrow, B,-1)$ <br>    $(B, \leftarrow, A,-1)$ <br>    $(A, \leftarrow$, exit, 10) |
| :--- |

Compute policy with

- $\gamma=1$
- estimate $\gamma$ which changes the policy computed for $\gamma=1$
- $\gamma=0$


## Example I

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10$)$ | $(D, \rightarrow$, exit, 6$)$ | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10$)$ |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

## Example I

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  | $(A, \leftarrow$, exit, 10) |  |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

## State set:

## Example I

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10$)$ | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10$)$ |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

State set: $S=\{A, B, C, D\}$, terminal states: $\{A, D\}$, non-terminal states: $\{B, C\}$

## Example I

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10$)$ | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10$)$ |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

State set: $S=\{A, B, C, D\}$, terminal states: $\{A, D\}$, non-terminal states: $\{B, C\}$ Action set:

## Example I

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10$)$ | $(D, \rightarrow$, exit, 6$)$ | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10$)$ |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

State set: $S=\{A, B, C, D\}$, terminal states: $\{A, D\}$, non-terminal states: $\{B, C\}$ Action set: $A=\{\rightarrow, \leftarrow\}$

## Example I

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10$)$ | $(D, \rightarrow$, exit, 6$)$ | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10$)$ |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

State set: $S=\{A, B, C, D\}$, terminal states: $\{A, D\}$, non-terminal states: $\{B, C\}$ Action set: $A=\{\rightarrow, \leftarrow\}$ Reward function:

## Example I

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10$)$ | $(D, \rightarrow$, exit, 6$)$ | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10$)$ |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

State set: $S=\{A, B, C, D\}$, terminal states: $\{A, D\}$, non-terminal states: $\{B, C\}$ Action set: $A=\{\rightarrow, \leftarrow\}$
Reward function: $r(\{B, C\})=-1, r(A)=10, r(D)=6$

## Example I

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10$)$ | $(D, \rightarrow$, exit, 6$)$ | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10$)$ |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

State set: $S=\{A, B, C, D\}$, terminal states: $\{A, D\}$, non-terminal states: $\{B, C\}$
Action set: $A=\{\rightarrow, \leftarrow\}$
Reward function: $r(\{B, C\})=-1, r(A)=10, r(D)=6$
Transition model:

## Example I

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10$)$ | $(D, \rightarrow$, exit, 6$)$ | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

State set: $S=\{A, B, C, D\}$, terminal states: $\{A, D\}$, non-terminal states: $\{B, C\}$
Action set: $A=\{\rightarrow, \leftarrow\}$
Reward function: $r(\{B, C\})=-1, r(A)=10, r(D)=6$
Transition model: $p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=2 / 2=1$

## Example I

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10$)$ | $(D, \rightarrow$, exit, 6$)$ | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

State set: $S=\{A, B, C, D\}$, terminal states: $\{A, D\}$, non-terminal states: $\{B, C\}$
Action set: $A=\{\rightarrow, \leftarrow\}$
Reward function: $r(\{B, C\})=-1, r(A)=10, r(D)=6$
Transition model: $p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=2 / 2=1$
World structure:

## Example I

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10$)$ | $(D, \rightarrow$, exit, 6$)$ | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

State set: $S=\{A, B, C, D\}$, terminal states: $\{A, D\}$, non-terminal states: $\{B, C\}$
Action set: $A=\{\rightarrow, \leftarrow\}$
Reward function: $r(\{B, C\})=-1, r(A)=10, r(D)=6$
Transition model: $p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=2 / 2=1$

World structure: | A | B | C | D |
| :--- | :--- | :--- | :--- |

## Example I

$$
\gamma=1
$$

## Example I, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10) |


| A | B | C | D |
| :--- | :--- | :--- | :--- |

$\mathrm{S}=\{A, B, C, D\}$
$A=\{\rightarrow, \leftarrow\}$
$r(\{B, C\})=-1, r(A)=10, r(D)=6$
$p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2=$
$p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1$
each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$
Estimate optimal policy:
A: $\pi(B)=\leftarrow, \pi(C)=\leftarrow$
B: $\pi(B)=\leftarrow, \pi(C)=\rightarrow$
C: $\pi(B)=\rightarrow, \pi(C)=\leftarrow$
D: $\pi(B)=\rightarrow, \pi(C)=\rightarrow$

## Example I, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10) |


| A | B | C | D |
| :--- | :--- | :--- | :--- |
| $\mathrm{S}=\{A, B, C, D\}$ |  |  |  |
| $A=\{\rightarrow, \leftarrow\}$ |  |  |  |

$r(\{B, C\})=-1, r(A)=10, r(D)=6$

| $p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2=$ |
| :--- |
| $p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1$ |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

Estimate optimal policy:
A: $\pi(B)=\leftarrow, \pi(C)=\leftarrow$
B: $\pi(B)=\leftarrow, \pi(C)=\rightarrow$
C: $\pi(B)=\rightarrow, \pi(C)=\leftarrow$
D: $\pi(B)=\rightarrow, \pi(C)=\rightarrow$
Let's find out :-)

## Example I, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

$$
\begin{aligned}
& \begin{array}{l}
\begin{array}{|c|c|c|c|}
\hline \mathrm{A} & \mathrm{~B} & \mathrm{C} & \mathrm{D} \\
\hline
\end{array} \quad \begin{array}{l}
S=\{A, B, C, D\}
\end{array}
\end{array} \\
& A=\{\rightarrow, \leftarrow\} \\
& r(\{B, C\})=-1, r(A)=10, r(D)=6 \\
& \begin{array}{l}
p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\
p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1
\end{array}
\end{aligned}
$$

## Compute:

A: $q(B, \leftarrow)=-1$
B: $q(B, \leftarrow)=5$
$C: q(B, \leftarrow)=9$
D: $q(B, \leftarrow)=6$

## Example I, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

## Compute:

A:
B:
C: $q(B, \leftarrow)=B \leftarrow A=10-1=9$
D:

$$
\begin{aligned}
& \begin{array}{l}
\begin{array}{|c|c|c|c|}
\hline \mathrm{A} & \mathrm{~B} & \mathrm{C} & \mathrm{D} \\
\hline \\
S=\{A, B, C, D\}
\end{array}
\end{array} \\
& A=\{\rightarrow, \leftarrow\} \\
& r(\{B, C\})=-1, r(A)=10, r(D)=6 \\
& \begin{array}{l}
p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\
p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1
\end{array}
\end{aligned}
$$

Example I, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  | $(A, \leftarrow$, exit, 10) |  |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

$$
-q(B, \leftarrow)=9
$$

## Compute:

A: $q(B, \rightarrow)=-1$
B: $q(B, \rightarrow)=7$
C: $q(B, \rightarrow)=10$
D: $q(B, \rightarrow)=6$

$$
\begin{aligned}
& \begin{array}{|l|l|l|l|}
\hline \mathrm{A} & \mathrm{~B} & \mathrm{C} & \mathrm{D} \\
\hline
\end{array} \\
& S=\{A, B, C, D\} \\
& A=\{\rightarrow, \leftarrow\} \\
& r(\{B, C\})=-1, r(A)=10, r(D)=6 \\
& \begin{array}{l}
p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\
p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1
\end{array}
\end{aligned}
$$

Example I, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

$$
-q(B, \leftarrow)=9
$$

## Compute:

A:
B: $q(B, \rightarrow)=B \rightarrow C \leftarrow B \leftarrow A=10-1-1-1=7$
C:
D:

$$
\begin{aligned}
& \begin{array}{|l|l|l|l|}
\hline \mathrm{A} & \mathrm{~B} & \mathrm{C} & \mathrm{D} \\
\hline
\end{array} \\
& S=\{A, B, C, D\} \\
& A=\{\rightarrow, \leftarrow\} \\
& r(\{B, C\})=-1, r(A)=10, r(D)=6 \\
& \begin{array}{l}
p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\
p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1
\end{array}
\end{aligned}
$$

## Example I, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

$$
\begin{aligned}
& \begin{array}{|l|l|l|l|}
\hline \mathrm{A} & \mathrm{~B} & \mathrm{C} & \mathrm{D} \\
\hline
\end{array} \\
& S=\{A, B, C, D\} \\
& A=\{\rightarrow, \leftarrow\} \\
& r(\{B, C\})=-1, r(A)=10, r(D)=6 \\
& \begin{array}{l}
p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\
p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1
\end{array}
\end{aligned}
$$

- $q(B, \leftarrow)=9$
- $q(B, \rightarrow)=7$

Compute:
A: $\pi(B)=\leftarrow$
B: $\pi(B)=\rightarrow$

## Example I, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  | $(A, \leftarrow$, exit, 10) |  |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

$$
\begin{aligned}
& \begin{array}{|l|l|l|l|}
\hline \mathrm{A} & \mathrm{~B} & \mathrm{C} & \mathrm{D} \\
\hline
\end{array} \\
& S=\{A, B, C, D\} \\
& A=\{\rightarrow, \leftarrow\} \\
& r(\{B, C\})=-1, r(A)=10, r(D)=6 \\
& \begin{array}{l}
p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\
p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1
\end{array}
\end{aligned}
$$

- $q(B, \leftarrow)=9$
- $q(B, \rightarrow)=7$

Compute:
A: $\pi(B)=\leftarrow$
B:

Example I, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

- $\pi(B)=\leftarrow$


## Compute:

A: $q(C, \rightarrow)=-1$
B: $q(C, \rightarrow)=5$
C: $q(C, \rightarrow)=9$
D: $q(C, \rightarrow)=6$

Example I, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

- $\pi(B)=\leftarrow$


## Compute:

A:
B: $q(C, \rightarrow)=C \rightarrow D=6-1=5$
C:

D:

Example I, $\gamma=1$


- $\pi(B)=\leftarrow$
$-q(C, \rightarrow)=5$
Compute:
A: $q(C, \leftarrow)=-1$
B: $q(C, \leftarrow)=6$
C: $q(C, \leftarrow)=10$
D: $q(C, \leftarrow)=8$

Example I, $\gamma=1$


- $\pi(B)=\leftarrow$
$-q(C, \rightarrow)=5$
Compute:
A:
B:
C:
D: $q(C, \leftarrow)=C \leftarrow B \leftarrow A=10-1-1=8$

Example I, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

| A | B | C | D |
| :---: | :---: | :---: | :---: |
| $S=\{A, B, C, D\}$ |  |  |  |
| $A=\{\rightarrow, \leftarrow\}$ |  |  |  |
| $r(\{B, C\})=-1, r(A)=10, r(D)=6$ |  |  |  |
| $\begin{aligned} & p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\ & p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1 \end{aligned}$ |  |  |  |

- $\pi(B)=\leftarrow$
- $q(C, \rightarrow)=5$
$-q(C, \leftarrow)=8$
Compute:
A: $\pi(C)=\leftarrow$
B: $\pi(C)=\rightarrow$

Example I, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

| A | B | C | D |
| :---: | :---: | :---: | :---: |
| $S=\{A, B, C, D\}$ |  |  |  |
| $A=\{\rightarrow, \leftarrow\}$ |  |  |  |
| $r(\{B, C\})=-1, r(A)=10, r(D)=6$ |  |  |  |
| $\begin{aligned} & p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\ & p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1 \end{aligned}$ |  |  |  |

- $\pi(B)=\leftarrow$
- $q(C, \rightarrow)=5$
$-q(C, \leftarrow)=8$
Compute:
A: $\pi(C)=\leftarrow$
B:


## Example I, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  | $(A, \leftarrow$, exit, 10) |  |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

$$
\begin{aligned}
& \begin{array}{l}
\begin{array}{|c|c|c|c|}
\hline \mathrm{A} & \mathrm{~B} & \mathrm{C} & \mathrm{D} \\
\hline \\
S=\{A, B, C, D\}
\end{array}
\end{array} \\
& A=\{\rightarrow, \leftarrow\} \\
& r(\{B, C\})=-1, r(A)=10, r(D)=6 \\
& \begin{array}{l}
p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\
p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1
\end{array}
\end{aligned}
$$

Evaluate policy for $\gamma=1$ :
A: $\pi(B)=\leftarrow, \pi(C)=\leftarrow$
B: $\pi(B)=\leftarrow, \pi(C)=\rightarrow$
$C: \pi(B)=\rightarrow, \pi(C)=\leftarrow$
D: $\pi(B)=\rightarrow, \pi(C)=\rightarrow$

## Example I, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

$$
\begin{aligned}
& \begin{array}{l}
\begin{array}{|c|c|c|c|}
\hline \mathrm{A} & \mathrm{~B} & \mathrm{C} & \mathrm{D} \\
\hline \\
S=\{A, B, C, D\}
\end{array}
\end{array} \\
& A=\{\rightarrow, \leftarrow\} \\
& r(\{B, C\})=-1, r(A)=10, r(D)=6 \\
& \begin{array}{l}
p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\
p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1
\end{array}
\end{aligned}
$$

Evaluate policy for $\gamma=1$ :
A: $\pi(B)=\leftarrow, \pi(C)=\leftarrow$
B:
C:
D:

## Example I

$$
\gamma=?
$$

## Example I, $\gamma=$ ?

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

| A | B | C | D |
| :---: | :---: | :---: | :---: |
| $S=\{A, B, C, D\}$ |  |  |  |
| $A=\{\rightarrow, \leftarrow\}$ |  |  |  |
| $r(\{B, C\})=-1, r(A)=10, r(D)=6$ |  |  |  |
| $\begin{aligned} & p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\ & p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1 \end{aligned}$ |  |  |  |

- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$


## Example I, $\gamma=$ ?



- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$


## Example I, $\gamma=$ ?

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10) |


| A | B | C | D |
| :---: | :---: | :---: | :---: |
| $S=\{A, B, C, D\}$ |  |  |  |
| $A=\{\rightarrow, \leftarrow\}$ |  |  |  |
| $r(\{B, C\})=-1, r(A)=10, r(D)=6$ |  |  |  |
| $\begin{aligned} & p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\ & p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1 \end{aligned}$ |  |  |  |

- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

How?
A: try some value and verify
B: compute boundary values
$C$ : guess

## Example I, $\gamma=$ ?

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10) |


| A | B | C | D |
| :---: | :---: | :---: | :---: |
| $S=\{A, B, C, D\}$ |  |  |  |
| $A=\{\rightarrow, \leftarrow\}$ |  |  |  |
| $r(\{B, C\})=-1, r(A)=10, r(D)=6$ |  |  |  |
| $\begin{aligned} & p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\ & p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1 \end{aligned}$ |  |  |  |

- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

How?
A: try some value and verify
B: compute boundary values
C:
: guess

## Example I, $\gamma=$ ?

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10) |


| A | B | C | D |
| :--- | :--- | :--- | :--- |
| $S=\{A, B, C, D\}$ |  |  |  |
| $A=\{\rightarrow, \leftarrow\}$ |  |  |  |

$r(\{B, C\})=-1, r(A)=10, r(D)=6$

| $p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2=$ |
| :--- |
| $p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1$ |

- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

Can the policy in state $B$ be changed?
A: Yes
B: No

Example I, $\gamma=$ ?

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10) |


| A | B | C | D |
| :--- | :--- | :--- | :--- |
| $S=\{A, B, C, D\}$ |  |  |  |
| $A=\{\rightarrow, \leftarrow\}$ |  |  |  |
| $r(\{B, C\})=-1, r(A)=10, r(D)=6$ |  |  |  |
| $p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2=$ |  |  |  |
| $p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1$ |  |  |  |

- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

Can the policy in state $B$ be changed?
A: Yes
B: No
Let's find out :-)

Example I, $\gamma=$ ?


- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

Can the policy in state $B$ be changed?

## Example I, $\gamma=$ ?



- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

Can the policy in state B be changed?
Compute:
A: $V(B)=\sum_{s^{\prime}} p\left(s^{\prime} \mid B, a\right)\left\{V\left(s^{\prime}\right)\right\}, s^{\prime} \in\{A, C\}$
B: $V(B)=\max _{a}\left(r(B)+\gamma \cdot V\left(s^{\prime}\right)\right), s^{\prime} \in\{A, C\}$
C: $V(B)=\arg \max _{a} \sum_{s^{\prime}} \gamma \cdot V\left(s^{\prime}\right), s^{\prime} \in\{A, C\}$
D: $V(B)=r(B)+\gamma V(D)$

## Example I, $\gamma=$ ?



- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

Can the policy in state $B$ be changed?
Compute:
A:
B: $V(B)=\max _{a}\left(r(B)+\gamma \cdot V\left(s^{\prime}\right)\right), s^{\prime} \in\{A, C\}$
C:
D:

## Example I, $\gamma=$ ?



- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

Can the policy in state $B$ be changed?
Compute:
A:
B: $V(B)=\max _{a}\left(r(B)+\gamma \cdot V\left(s^{\prime}\right)\right), s^{\prime} \in\{A, C\}$
C:

D:
$\Rightarrow$ depends on $V(A), V(C)$

## Example I, $\gamma=$ ?

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$


- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

Can the policy in state B be changed?
Determine:
A: $V(A)<V(C)$
B: $V(A)=V(C)$
C: $V(A)>V(C)$

Example I, $\gamma=$ ?


- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

Can the policy in state $B$ be changed?
Determine:
A:
B:
C: $V(A)>V(C) ; V(A)=10, V(C)<10$

## Example I, $\gamma=$ ?



- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

Can the policy in state $B$ be changed?
Determine:
A:
B:
C: $V(A)>V(C) ; V(A)=10, V(C)<10$

$$
\begin{aligned}
\Rightarrow & \pi(B)=\leftarrow \\
& V(B)=r(B)+\gamma \cdot V(A)=-1+10 \gamma
\end{aligned}
$$

Example I, $\gamma=$ ?


- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

State $B: \pi(B)=\leftarrow, V(B)=-1+10 \gamma$

## Example I, $\gamma=$ ?



- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

State $B: \pi(B)=\leftarrow, V(B)=-1+10 \gamma$ $\Rightarrow$ Policy in state $C$ has to be changed.

## Example I, $\gamma=$ ?



- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

State $B: \pi(B)=\leftarrow, V(B)=-1+10 \gamma$
$\Rightarrow$ Policy in state $C$ has to be changed.
How?
A: $q(C, \rightarrow)>q(C, \leftarrow)$
B: $V(C)>V(B)$
C: $\pi(C)>\pi(B)$
D: $r(C)>r(B)$

## Example I, $\gamma=$ ?



- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

State $B: \pi(B)=\leftarrow, V(B)=-1+10 \gamma$
$\Rightarrow$ Policy in state $C$ has to be changed.
How?
A: $q(C, \rightarrow)>q(C, \leftarrow)$
B:
C:
D:

## Example I, $\gamma=$ ?



- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

State $B: \pi(B)=\leftarrow, V(B)=-1+10 \gamma$ $\Rightarrow$ Policy in state $C$ has to be changed.
Compute:
A: $q(C, \rightarrow)=r(C)+V(D)$
B: $q(C, \rightarrow)=r(C)+\gamma \cdot V(B)$
C: $q(C, \rightarrow)=r(C)+\gamma \cdot V(D)$
D: $q(C, \rightarrow)=r(C)+V(B)$

## Example I, $\gamma=$ ?



- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

State $B: \pi(B)=\leftarrow, V(B)=-1+10 \gamma$ $\Rightarrow$ Policy in state $C$ has to be changed.
Compute:
A:
B:
C: $q(C, \rightarrow)=r(C)+\gamma \cdot V(D)=-1+6 \gamma$
D:

## Example I, $\gamma=$ ?



- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

State $B: \pi(B)=\leftarrow, V(B)=-1+10 \gamma$ $\Rightarrow$ Policy in state $C$ has to be changed.

- $q(C, \rightarrow)=-1+6 \gamma$

Compute:
A: $q(C, \leftarrow)=r(C)+V(D)$
B: $q(C, \leftarrow)=r(C)+\gamma \cdot V(B)$
C: $q(C, \leftarrow)=r(C)+\gamma \cdot V(D)$
$\mathrm{D}: q(C, \leftarrow)=r(C)+V(B)$

Example I, $\gamma=$ ?


- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

State $B: \pi(B)=\leftarrow, V(B)=-1+10 \gamma$ $\Rightarrow$ Policy in state $C$ has to be changed.

- $q(C, \rightarrow)=-1+6 \gamma$

Compute:
A:
B: $q(C, \leftarrow)=r(C)+\gamma \cdot V(B)=-1+\gamma(-1+10 \gamma)$
C:
D:

Example I, $\gamma=$ ?


- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

State $B: \pi(B)=\leftarrow, V(B)=-1+10 \gamma$
$\Rightarrow$ Policy in state $C$ has to be changed.
$-q(C, \rightarrow)=-1+6 \gamma$

- $q(C, \leftarrow)=-1+\gamma(-1+10 \gamma)$

To change the policy, we need:
$q(C, \rightarrow)>q(C, \leftarrow)$

Example I, $\gamma=$ ?


- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

$$
q(C, \rightarrow)>q(C, \leftarrow)
$$

Example I, $\gamma=$ ?


- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

$$
\begin{aligned}
q(C, \rightarrow) & >q(C, \leftarrow) \\
-1+6 \gamma & >-1+\gamma(-1+10 \gamma)
\end{aligned}
$$

Example I, $\gamma=$ ?


- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

$$
\begin{aligned}
q(C, \rightarrow) & >q(C, \leftarrow) \\
-1+6 \gamma & >-1+\gamma(-1+10 \gamma) \\
-1+6 \gamma & >-1-\gamma+10 \gamma^{2}
\end{aligned}
$$

Example I, $\gamma=$ ?


- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

$$
\begin{aligned}
q(C, \rightarrow) & >q(C, \leftarrow) \\
-1+6 \gamma & >-1+\gamma(-1+10 \gamma) \\
-1+6 \gamma & >-1-\gamma+10 \gamma^{2} \\
7 \gamma-10 \gamma^{2} & >0
\end{aligned}
$$

## Example I, $\gamma=$ ?



- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

$$
\begin{aligned}
q(C, \rightarrow) & >q(C, \leftarrow) \\
-1+6 \gamma & >-1+\gamma(-1+10 \gamma) \\
-1+6 \gamma & >-1-\gamma+10 \gamma^{2} \\
7 \gamma-10 \gamma^{2} & >0 \\
\Rightarrow \gamma_{1}=0, \gamma_{2} & =0.7
\end{aligned}
$$

## Example I, $\gamma=$ ?



- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

$$
\begin{aligned}
& q(C, \rightarrow)>q(C, \leftarrow) \\
&-1+6 \gamma>-1+\gamma(-1+10 \gamma) \\
&-1+6 \gamma>-1-\gamma+10 \gamma^{2} \\
& 7 \gamma-10 \gamma^{2}>0 \\
& \Rightarrow \gamma_{1}=0, \gamma_{2}=0.7 \\
&\Rightarrow \pi(B)=\leftarrow, \pi(C)=\leftarrow ; \text { for } \gamma \in] 0.7,1] \\
&\pi(B)=\leftarrow, \pi(C)=\rightarrow \text {; for } \gamma \in] 0,0.7[
\end{aligned}
$$

## Example I

$$
\gamma=0
$$

## Example I, $\gamma=0$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  | $(A, \leftarrow$, exit, 10) |  |

Foach field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$
For $\gamma=0$ :

$$
\begin{aligned}
& \begin{array}{|l|l|l|l|}
\hline \mathrm{A} & \mathrm{~B} & \mathrm{C} & \mathrm{D} \\
\hline
\end{array} \\
& S=\{A, B, C, D\} \\
& A=\{\rightarrow, \leftarrow\} \\
& r(\{B, C\})=-1, r(A)=10, r(D)=6 \\
& \begin{array}{l}
p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\
p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1
\end{array}
\end{aligned}
$$

Compute:
A: $V(B)=9$
B: $V(B)=5$
C: $V(B)=-1$
D: $V(B)=0$

## Example I, $\gamma=0$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  | $(A, \leftarrow$, exit, 10) |  |

Foach field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$
For $\gamma=0$ :

$$
\begin{aligned}
& \begin{array}{|l|l|l|l|}
\hline \mathrm{A} & \mathrm{~B} & \mathrm{C} & \mathrm{D} \\
\hline
\end{array} \\
& S=\{A, B, C, D\} \\
& A=\{\rightarrow, \leftarrow\} \\
& r(\{B, C\})=-1, r(A)=10, r(D)=6 \\
& \begin{array}{l}
p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\
p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1
\end{array}
\end{aligned}
$$

Compute:

A:

B:

C: $V(B)=r(B)+\gamma \max _{a} V\left(s^{\prime}\right)=-1+0=-1$
D:

Example I, $\gamma=0$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10) | $(D, \rightarrow$, exit,6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10) |

each field in the table is an n-tuple ( $s, a, s^{\prime}, r$ )
For $\gamma=0$ :

- $V(B)=-1$ for $\{\leftarrow, \rightarrow\}$

Compute:
A: $V(C)=9$
B: $V(C)=5$
C: $V(C)=-1$
D: $V(C)=0$

| A |
| :--- |
| B |
| B |
| C |
| C |
| D |
| $\mathrm{S}=\{A, B, C, D\}$ |
| $A=\{\rightarrow, \leftarrow\}$ |
| $r(\{B, C\})=-1, r(A)=10, r(D)=6$ |
| $p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2=$ <br> $p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1$ |

$$
S=\{A, B, C, D\}
$$

$$
A=\{\rightarrow, \leftarrow\}
$$

$$
r(\{B, C\})=-1, r(A)=10, r(D)=6
$$

$$
\begin{aligned}
& p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\
& p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1
\end{aligned}
$$

Example I, $\gamma=0$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10) | $(D, \rightarrow$, exit,6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$
For $\gamma=0$ :

- $V(B)=-1$ for $\{\leftarrow, \rightarrow\}$

Compute:
A:
$B$ :
C: $V(C)=r(C)+\gamma \max _{a} V\left(s^{\prime}\right)=-1+0=-1$
D:


$$
\frac{\mathrm{A}|\mathrm{~B}| \mathrm{C} \mid}{S=\{A, B, C, D\}}
$$

$$
A=\{\rightarrow, \leftarrow\}
$$

$$
r(\{B, C\})=-1, r(A)=10, r(D)=6
$$

$$
\begin{aligned}
& p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\
& p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1
\end{aligned}
$$

## Example I, $\gamma=0$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10 $)$ | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10) |


| A | B | C | D |
| :--- | :--- | :--- | :--- |
| $S=\{A, B, C, D\}$ |  |  |  |
| $A=\{\rightarrow, \leftarrow\}$ |  |  |  |
| $r(\{B, C\})=-1, r(A)=10, r(D)=6$ |  |  |  |
| $p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2=$ |  |  |  |
| $p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1$ |  |  |  | For $\gamma=0$ :

- $V(B)=-1$ for $\{\leftarrow, \rightarrow\}$
- $V(C)=-1$ for $\{\leftarrow, \rightarrow\}$


## Example I, $\gamma=0$

| Episode 1 | Episode 2 |
| :---: | :---: |
| $\begin{aligned} & (B, \rightarrow, C,-1) \\ & (C, \rightarrow, D,-1) \\ & (D, \leftarrow, \text { exit }, 6) \end{aligned}$ | $\begin{gathered} (B, \leftarrow, A,-1) \\ (A, \rightarrow, \text { exit }, 10) \end{gathered}$ |
| each field in the table is an n-tuple ( For $\gamma=0$ : |  |
| - $V(B)=-1$ for $\{\leftarrow, \rightarrow\}$ |  |
| $\begin{aligned} \Rightarrow \quad \pi(B) & =\{\leftarrow \\ \pi(C) & =\{\psi \end{aligned}$ |  |


| A | B | C | D |
| :--- | :--- | :--- | :--- |
| $S=\{A, B, C, D\}$ |  |  |  |
| $A=\{\rightarrow, \leftarrow\}$ |  |  |  |
| $r(\{B, C\})=-1, r(A)=10, r(D)=6$ |  |  |  |
| $p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2=$ |  |  |  |
| $p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1$ |  |  |  |

## Example I

summary

Example I, summary

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-1)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-1)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-1)$ | $(A, \rightarrow$, exit, 10) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-1)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 10) |

$$
\begin{aligned}
& \hline \mathrm{A} \\
& \hline \mathrm{~B} \\
& \hline
\end{aligned} \mathrm{C}|\mathrm{D}| \mathrm{C}, ~ \begin{aligned}
& \mathrm{S}=\{A, B, C, D\} \\
& A=\{\rightarrow, \leftarrow\} \\
& r(\{B, C\})=-1, r(A)=10, r(D)=6 \\
& \begin{array}{l}
p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\
p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1
\end{array}
\end{aligned}
$$

For $\gamma=1: \quad \pi(B)=\leftarrow$

$$
\pi(C)=\leftarrow
$$

For $\gamma \in] 0.7,1]: \quad \pi(B)=\leftarrow$

$$
\pi(C)=\leftarrow
$$

For $\gamma \in] 0,0.7[: \quad \pi(B)=\leftarrow$

$$
\pi(C)=\rightarrow
$$

For $\gamma=0$ :

$$
\begin{aligned}
& \pi(B)=\{\leftarrow, \rightarrow\} \\
& \pi(C)=\{\leftarrow, \rightarrow\}
\end{aligned}
$$

## Example II

## Example II

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

## Example II

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit,6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

Compute policy with

- $\gamma=1$
- estimate $\gamma$ that changes the policy for $\gamma=1$
- $\gamma=0$


## Example II

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit,6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit,6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

## Example II

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

State set:

## Example II

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  | $(C, \leftarrow, B,-1)$ |  |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

State set: $S=\{A, B, C, D\}$, terminal states: $\{A, D\}$, non-terminal states: $\{B, C\}$

## Example II

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  | $(C, \leftarrow, B,-1)$ |  |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

State set: $S=\{A, B, C, D\}$, terminal states: $\{A, D\}$, non-terminal states: $\{B, C\}$ Action set:

## Example II

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  | $(C, \leftarrow, B,-1)$ |  |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

State set: $S=\{A, B, C, D\}$, terminal states: $\{A, D\}$, non-terminal states: $\{B, C\}$ Action set: $A=\{\rightarrow, \leftarrow\}$

## Example II

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  | $(A, \leftarrow$, exit, 6) |  |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

State set: $S=\{A, B, C, D\}$, terminal states: $\{A, D\}$, non-terminal states: $\{B, C\}$ Action set: $A=\{\rightarrow, \leftarrow\}$
Reward function:

## Example II

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  | $(C, \leftarrow, B,-1)$ |  |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

State set: $S=\{A, B, C, D\}$, terminal states: $\{A, D\}$, non-terminal states: $\{B, C\}$ Action set: $A=\{\rightarrow, \leftarrow\}$
Reward function: $r(\{B, C\}, \leftarrow)=-1, r(\{B, C\}, \rightarrow)=-3, r(\{A, D\})=6$

## Example II

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  | $(C, \leftarrow, B,-1)$ |  |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

State set: $S=\{A, B, C, D\}$, terminal states: $\{A, D\}$, non-terminal states: $\{B, C\}$
Action set: $A=\{\rightarrow, \leftarrow\}$
Reward function: $r(\{B, C\}, \leftarrow)=-1, r(\{B, C\}, \rightarrow)=-3, r(\{A, D\})=6$
Transition model:

## Example II

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit,6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit,6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

State set: $S=\{A, B, C, D\}$, terminal states: $\{A, D\}$, non-terminal states: $\{B, C\}$
Action set: $A=\{\rightarrow, \leftarrow\}$
Reward function: $r(\{B, C\}, \leftarrow)=-1, r(\{B, C\}, \rightarrow)=-3, r(\{A, D\})=6$
Transition model: $p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=2 / 2=1$

## Example II

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

State set: $S=\{A, B, C, D\}$, terminal states: $\{A, D\}$, non-terminal states: $\{B, C\}$
Action set: $A=\{\rightarrow, \leftarrow\}$
Reward function: $r(\{B, C\}, \leftarrow)=-1, r(\{B, C\}, \rightarrow)=-3, r(\{A, D\})=6$
Transition model: $p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=2 / 2=1$
World structure:

## Example II

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

State set: $S=\{A, B, C, D\}$, terminal states: $\{A, D\}$, non-terminal states: $\{B, C\}$
Action set: $A=\{\rightarrow, \leftarrow\}$
Reward function: $r(\{B, C\}, \leftarrow)=-1, r(\{B, C\}, \rightarrow)=-3, r(\{A, D\})=6$
Transition model: $p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=2 / 2=1$
World structure:

| A | B | C | D |
| :--- | :--- | :--- | :--- |

Example II

$$
\gamma=1
$$

## Example II, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

$$
\begin{aligned}
& \begin{array}{|c|c|c|c|}
\hline \mathrm{A} & \mathrm{~B} & \mathrm{C} & \mathrm{D} \\
\hline S=\{A, B, C, D\} \\
A=\{\rightarrow, \leftarrow\} \\
r(\{B, C\}, \leftarrow)=-1, r(\{A, D\})=6, \\
r(\{B, C\}, \rightarrow)=-3 \\
p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\
p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1
\end{array}
\end{aligned}
$$

Estimate optimal policy:
A: $\pi(B)=\leftarrow, \pi(C)=\leftarrow$
B: $\pi(B)=\leftarrow, \pi(C)=\rightarrow$
C: $\pi(B)=\rightarrow, \pi(C)=\leftarrow$
D: $\pi(B)=\rightarrow, \pi(C)=\rightarrow$

## Example II, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

$$
\begin{aligned}
& \begin{array}{|c|c|c|c|}
\hline \mathrm{A} & \mathrm{~B} & \mathrm{C} & \mathrm{D} \\
\hline S=\{A, B, C, D\} \\
A=\{\rightarrow, \leftarrow\} \\
r(\{B, C\}, \leftarrow)=-1, r(\{A, D\})=6, \\
r(\{B, C\}, \rightarrow)=-3 \\
p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\
p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1
\end{array}
\end{aligned}
$$

Estimate optimal policy:
A: $\pi(B)=\leftarrow, \pi(C)=\leftarrow$
B: $\pi(B)=\leftarrow, \pi(C)=\rightarrow$
C: $\pi(B)=\rightarrow, \pi(C)=\leftarrow$
D: $\pi(B)=\rightarrow, \pi(C)=\rightarrow$
Let's find out :-)

## Example II, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  | $(A, \leftarrow$, exit, 6) |  |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

$$
\begin{aligned}
& \begin{array}{|c|c|c|c|}
\hline \mathrm{A} & \mathrm{~B} & \mathrm{C} & \mathrm{D} \\
\hline S=\{A, B, C, D\} \\
A=\{\rightarrow, \leftarrow\} \\
r(\{B, C\}, \leftarrow)=-1, r(\{A, D\})=6, \\
r(\{B, C\}, \rightarrow)=-3 \\
p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\
p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1
\end{array}
\end{aligned}
$$

Compute:
A: $q(B, \leftarrow)=-1$
B: $q(B, \leftarrow)=5$
C: $q(B, \leftarrow)=-3$
D: $q(B, \leftarrow)=6$

## Example II, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  | $(A, \leftarrow$, exit, 6) |  |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

$$
\begin{aligned}
& \begin{array}{|c|c|c|c|}
\hline \mathrm{A} & \mathrm{~B} & \mathrm{C} & \mathrm{D} \\
\hline S=\{A, B, C, D\} \\
A=\{\rightarrow, \leftarrow\} \\
r(\{B, C\}, \leftarrow)=-1, r(\{A, D\})=6, \\
r(\{B, C\}, \rightarrow)=-3 \\
p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\
p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1
\end{array}
\end{aligned}
$$

Compute:

## A:

B: $q(B, \leftarrow)=B \leftarrow A=6-1=5$
C:
D:

Example II, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  | $(A, \leftarrow$, exit, 6) |  |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$
$-q(B, \leftarrow)=5$
Compute:
A: $q(B, \rightarrow)=-1$
B: $q(B, \rightarrow)=0$
C: $q(B, \rightarrow)=1$
D: $q(B, \rightarrow)=-3$

| A | B | C | D |
| :--- | :--- | :--- | :--- |

$$
S=\{A, B, C, D\}
$$

$$
A=\{\rightarrow, \leftarrow\}
$$

$$
r(\{B, C\}, \leftarrow)=-1, \quad r(\{A, D\})=6
$$

$$
r(\{B, C\}, \rightarrow)=-3
$$

$$
\begin{aligned}
& p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\
& p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1
\end{aligned}
$$

Example II, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  | $(A, \leftarrow$, exit, 6) |  |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

$$
\begin{aligned}
& \begin{array}{|c|c|c|c|}
\hline \mathrm{A} & \mathrm{~B} & \mathrm{C} & \mathrm{D} \\
\hline S=\{A, B, C, D\} \\
A=\{\rightarrow, \leftarrow\} \\
r(\{B, C\}, \leftarrow)=-1, r(\{A, D\})=6, \\
r(\{B, C\}, \rightarrow)=-3 \\
p(C \mid B, \rightarrow)=p_{p}(A \mid B, \leftarrow)=2 / 2= \\
p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1
\end{array}
\end{aligned}
$$

$-q(B, \leftarrow)=5$
Compute:

A:

B:
$\mathrm{C}: q(B, \rightarrow)=B \rightarrow C \leftarrow B \leftarrow A=6-3-1-1=1$

D:

## Example II, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  | $(A, \leftarrow$, exit, 6) |  |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

| A | B | C | D |
| :--- | :--- | :--- | :--- |
| $S=\{A, B, C, D\}$ |  |  |  |
| $A=\{\rightarrow, \leftarrow\}$ |  |  |  |
| $r(\{B, C\}, \leftarrow)=-1, \quad r(\{A, D\})=6$, |  |  |  |
| $r(\{B, C\}, \rightarrow)=-3$ |  |  |  |
| $p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2=$ |  |  |  |
| $p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1$ |  |  |  |

$-q(B, \leftarrow)=5$

- $q(B, \rightarrow)=1$

Compute:
A: $\pi(B)=\leftarrow$
B: $\pi(B)=\rightarrow$

## Example II, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  | $(A, \leftarrow$, exit, 6) |  |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

| A | B | C | D |
| :--- | :--- | :--- | :--- |
| $S=\{A, B, C, D\}$ |  |  |  |
| $A=\{\rightarrow, \leftarrow\}$ |  |  |  |
| $r(\{B, C\}, \leftarrow)=-1, \quad r(\{A, D\})=6$, |  |  |  |
| $r(\{B, C\}, \rightarrow)=-3$ |  |  |  |
| $p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2=$ |  |  |  |
| $p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1$ |  |  |  |

$-q(B, \leftarrow)=5$

- $q(B, \rightarrow)=1$

Compute:
A: $\pi(B)=\leftarrow$
B:

Example II, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

- $\pi(B)=\leftarrow$

Compute:
A: $q(C, \rightarrow)=-1$
B: $q(C, \rightarrow)=-3$
C: $q(C, \rightarrow)=3$
D: $q(C, \rightarrow)=6$

$$
\begin{aligned}
& \begin{array}{|c|c|c|c|}
\hline \mathrm{A} & \mathrm{~B} & \mathrm{C} & \mathrm{D} \\
\hline S=\{A, B, C, D\} \\
A=\{\rightarrow, \leftarrow\} \\
r(\{B, C\}, \leftarrow)=-1, \quad r(\{A, D\})=6, \\
r(\{B, C\}, \rightarrow)=-3 \\
p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\
p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1
\end{array}
\end{aligned}
$$

Example II, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

- $\pi(B)=\leftarrow$

Compute:
A:

B:
C: $q(C, \rightarrow)=C \rightarrow D=6-3=3$
D:

$$
\begin{aligned}
& \begin{array}{|l|l|l|l|}
\hline \mathrm{A} & \mathrm{~B} & \mathrm{C} & \mathrm{D} \\
\hline
\end{array} \\
& S=\{A, B, C, D\} \\
& A=\{\rightarrow, \leftarrow\} \\
& r(\{B, C\}, \leftarrow)=-1, r(\{A, D\})=6 \text {, } \\
& r(\{B, C\}, \rightarrow)=-3 \\
& \begin{array}{l}
p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\
p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1
\end{array}
\end{aligned}
$$

Example II, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

- $\pi(B)=\leftarrow$
- $q(C, \rightarrow)=3$


## Compute:

A: $q(C, \leftarrow)=-1$
B: $q(C, \leftarrow)=6$
C: $q(C, \leftarrow)=-3$
D: $q(C, \leftarrow)=4$

Example II, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

- $\pi(B)=\leftarrow$
$-q(C, \rightarrow)=3$


## Compute: <br> Comput

A:
B:
C:
D: $q(C, \leftarrow)=C \leftarrow B \leftarrow A=6-1-1=4$
A:

| A | B | C | D |
| :--- | :--- | :--- | :--- |

$\mathrm{S}=\{A, B, C, D\}$
$A=\{\rightarrow, \leftarrow\}$
$r(\{B, C\}, \leftarrow)=-1, \quad r(\{A, D\})=6$,
$r(\{B, C\}, \rightarrow)=-3$
$p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2=$
$p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1$

$$
\begin{aligned}
& A|B| C \mid \\
& S=\{A, B, C, D\}
\end{aligned}
$$

$$
A=\{\rightarrow, \leftarrow\}
$$

$$
r(\{B, C\}, \leftarrow)=-1, \quad r(\{A, D\})=6,
$$

$$
r(\{B, C\}, \rightarrow)=-3
$$

$$
\begin{aligned}
& p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\
& p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1
\end{aligned}
$$

Example II, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

- $\pi(B)=\leftarrow$
- $q(C, \rightarrow)=3$
$-q(C, \leftarrow)=4$
Compute:
A: $\pi(C)=\leftarrow$
B: $\pi(C)=\rightarrow$

| A | B | C | D |
| :---: | :---: | :---: | :---: |
| $S=\{A, B, C, D\}$ |  |  |  |
| $A=\{\rightarrow, \leftarrow\}$ |  |  |  |
| $\begin{aligned} & r(\{B, C\}, \leftarrow)=- \\ & r(\{B, C\}, \rightarrow)=-3 \end{aligned}$ |  |  |  |
| $\begin{aligned} & p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\ & p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1 \end{aligned}$ |  |  |  |

Example II, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

- $\pi(B)=\leftarrow$
- $q(C, \rightarrow)=3$
$-q(C, \leftarrow)=4$
Compute:
A: $\pi(C)=\leftarrow$
B:

| A | B | C | D |
| :--- | :--- | :--- | :--- |

$\mathrm{S}=\{A, B, C, D\}$
$A=\{\rightarrow, \leftarrow\}$
$r(\{B, C\}, \leftarrow)=-1, \quad r(\{A, D\})=6$,
$r(\{B, C\}, \rightarrow)=-3$
$p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2=$
$p(D \mid C) \rightarrow p(B \mid C, \leftarrow)=1$

## Example II, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  | $(A, \leftarrow$, exit, 6) |  |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

$$
\begin{aligned}
& \hline \mathrm{A} \mid \mathrm{B} \\
& \hline \mathrm{C} \\
& \hline \mathrm{D} \\
& \begin{array}{l}
\mathrm{S}=\{A, B, C, D\} \\
A=\{\rightarrow, \leftarrow\} \\
r(\{B, C\}, \leftarrow)=-1, \quad r(\{A, D\})=6, \\
r(\{B, C\}, \rightarrow)=-3 \\
p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\
p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1
\end{array}
\end{aligned}
$$

Evaluate policy for $\gamma=1$ :
A: $\pi(B)=\leftarrow, \pi(C)=\leftarrow$
B: $\pi(B)=\leftarrow, \pi(C)=\rightarrow$
C: $\pi(B)=\rightarrow, \pi(C)=\leftarrow$
D: $\pi(B)=\rightarrow, \pi(C)=\rightarrow$

## Example II, $\gamma=1$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  | $(A, \leftarrow$, exit, 6) |  |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

$$
\begin{aligned}
& \begin{array}{|c|c|c|c|}
\hline \mathrm{A} & \mathrm{~B} & \mathrm{C} & \mathrm{D} \\
\hline S=\{A, B, C, D\} \\
A=\{\rightarrow, \leftarrow\} \\
r(\{B, C\}, \leftarrow)=-1, r(\{A, D\})=6, \\
r(\{B, C\}, \rightarrow)=-3 \\
p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\
p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1
\end{array}
\end{aligned}
$$

Evaluate policy for $\gamma=1$ :
A: $\pi(B)=\leftarrow, \pi(C)=\leftarrow$
B:
C:
D:

Example II

$$
\gamma=?
$$

## Example II, $\gamma=$ ?

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

Can the policy in state $B$ be changed?
A: Yes
B: No

## Example II, $\gamma=$ ?

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |


| A | B | C | D |
| :--- | :--- | :--- | :--- |
| $S=\{A, B, C, D\}$ |  |  |  |
| $A=\{\rightarrow, \leftarrow\}$ |  |  |  |
| $r(\{B, C\}, \leftarrow)=-1, \quad r(\{A, D\})=6$, |  |  |  |
| $r(\{B, C\}, \rightarrow)=-3$ |  |  |  |
| $p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2=$ |  |  |  |
| $p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1$ |  |  |  |

- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

Can the policy in state $B$ be changed?
A: Yes
B: No
Let's find out :-)

Example II, $\gamma=$ ?

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

Can the policy in state $B$ be changed?

Example II, $\gamma=$ ?

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

Can the policy in state B be changed?
$\left.V(B)=\max _{a}\left(r(B, a)+\gamma \cdot V\left(s^{\prime}\right)\right), s^{\prime} \in\{A, C\}\right)$

Example II, $\gamma=$ ?

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

Can the policy in state B be changed?

$$
\begin{aligned}
V(B) & \left.=\max _{a}\left(r(B, a)+\gamma \cdot V\left(s^{\prime}\right)\right), s^{\prime} \in\{A, C\}\right) \\
& =\max \left\{\begin{array}{ll}
(\rightarrow) & r(B, \rightarrow)+\gamma \cdot V(C) \\
(\leftarrow) & r(B, \leftarrow)+\gamma \cdot V(A)
\end{array}\right\}
\end{aligned}
$$

Example II, $\gamma=$ ?

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

Can the policy in state B be changed?

$$
\begin{aligned}
V(B) & \left.=\max _{a}\left(r(B, a)+\gamma \cdot V\left(s^{\prime}\right)\right), s^{\prime} \in\{A, C\}\right) \\
& =\max \left\{\begin{array}{lr}
(\rightarrow) & r(B, \rightarrow)+\gamma \cdot V(C) \\
(\leftarrow) r(B, \leftarrow)+\gamma \cdot V(A)
\end{array}\right\} \\
& =r(B, \leftarrow)+\gamma \cdot V(A) \text { since } V(A)>V(C) \& r(B, \leftarrow)>r(B, \rightarrow)
\end{aligned}
$$

Example II, $\gamma=$ ?

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$


- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

Can the policy in state B be changed?

$$
\begin{aligned}
V(B) & \left.=\max _{a}\left(r(B, a)+\gamma \cdot V\left(s^{\prime}\right)\right), s^{\prime} \in\{A, C\}\right) \\
& =\max \left\{\begin{array}{ll}
(\rightarrow) & r(B, \rightarrow)+\gamma \cdot V(C) \\
(\leftarrow) & r(B, \leftarrow)+\gamma \cdot V(A)
\end{array}\right\} \\
& =r(B, \leftarrow)+\gamma \cdot V(A) \text { since } V(A)>V(C) \& r(B, \leftarrow)>r(B, \rightarrow) \\
\Rightarrow \pi(B) & =\leftarrow \\
V(B) & =-1+6 \gamma
\end{aligned}
$$

Example II, $\gamma=$ ?

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

State $B: \pi(B)=\leftarrow, V(B)=-1+6 \gamma$

Example II, $\gamma=$ ?

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

State $B: \pi(B)=\leftarrow, V(B)=-1+6 \gamma$
$\Rightarrow$ Policy in state $C$ has to be changed.

Example II, $\gamma=$ ?

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$
$\checkmark$ for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$

- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

State $B: \pi(B)=\leftarrow, V(B)=-1+6 \gamma$
$\Rightarrow$ Policy in state $C$ has to be changed.

## Compute:

A: $q(C, \rightarrow)=r(C, \rightarrow)+\gamma \cdot V(D)$
B: $q(C, \rightarrow)=r(C, \rightarrow)+\gamma \cdot V(B)$
C: $q(C, \rightarrow)=r(C, \leftarrow)+\gamma \cdot V(D)$
D: $q(C, \rightarrow)=r(C, \leftarrow)+V(B)$

Example II, $\gamma=$ ?

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$
$\checkmark$ for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$

- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

State $B: \pi(B)=\leftarrow, V(B)=-1+6 \gamma$
$\Rightarrow$ Policy in state $C$ has to be changed.

## Compute:

A: $q(C, \rightarrow)=r(C, \rightarrow)+\gamma \cdot V(D)=-3+6 \gamma$
B:
C:
D:

Example II, $\gamma=$ ?

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$


- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

State $B: \pi(B)=\leftarrow, V(B)=-1+6 \gamma$
$\Rightarrow$ Policy in state $C$ has to be changed.
$-q(C, \rightarrow)=-3+6 \gamma$
Compute:
A: $q(C, \leftarrow)=r(C)+V(D)$
B: $q(C, \leftarrow)=r(C, \leftarrow)+\gamma \cdot V(B)$
$\mathrm{C}: q(C, \leftarrow)=r(C \leftarrow)+\gamma \cdot V(D)$
$\mathrm{D}: q(C, \leftarrow)=r(C, \rightarrow)+V(B)$

Example II, $\gamma=$ ?

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$


- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

State $B: \pi(B)=\leftarrow, V(B)=-1+6 \gamma$
$\Rightarrow$ Policy in state $C$ has to be changed.
$-q(C, \rightarrow)=-3+6 \gamma$
Compute:
A:
B: $q(C, \leftarrow)=r(C, \leftarrow)+\gamma \cdot V(B)=-1+\gamma(-1+6 \gamma)$
C:
D:

Example II, $\gamma=$ ?

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple ( $s, a, s^{\prime}, r$ )


- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

State $B: \pi(B)=\leftarrow, V(B)=-1+6 \gamma$
$\Rightarrow$ Policy in state $C$ has to be changed.
$-q(C, \rightarrow)=-3+6 \gamma$
$-q(C, \leftarrow)=-1+\gamma(-1+6 \gamma)$
To change the policy, we need: $q(C, \rightarrow)>q(C, \leftarrow)$

Example II, $\gamma=$ ?

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple ( $s, a, s^{\prime}, r$ )


- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

State $B: \pi(B)=\leftarrow, V(B)=-1+6 \gamma$
$\Rightarrow$ Policy in state $C$ has to be changed.
$-q(C, \rightarrow)=-3+6 \gamma$
$-q(C, \leftarrow)=-1+\gamma(-1+6 \gamma)$
To change the policy, we need:
$q(C, \rightarrow)>q(C, \leftarrow)$
Let's evaluate for boundary equality:
$q(C, \rightarrow)=q(C, \leftarrow)$

Example II, $\gamma=$ ?

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple ( $s, a, s^{\prime}, r$ )


- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

$$
q(C, \rightarrow)=q(C, \leftarrow)
$$

Example II, $\gamma=$ ?

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit,6) |

each field in the table is an n-tuple ( $s, a, s^{\prime}, r$ )


- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

$$
\begin{aligned}
q(C, \rightarrow) & =q(C, \leftarrow) \\
-3+6 \gamma & =-1+\gamma(-1+6 \gamma)
\end{aligned}
$$

Example II, $\gamma=$ ?

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit,6) |

each field in the table is an n-tuple ( $s, a, s^{\prime}, r$ )


- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

$$
\begin{aligned}
q(C, \rightarrow) & =q(C, \leftarrow) \\
-3+6 \gamma & =-1+\gamma(-1+6 \gamma) \\
-3+6 \gamma & =-1-\gamma+6 \gamma^{2}
\end{aligned}
$$

Example II, $\gamma=$ ?

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit,6) |

each field in the table is an n-tuple ( $s, a, s^{\prime}, r$ )


- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

$$
\begin{aligned}
q(C, \rightarrow) & =q(C, \leftarrow) \\
-3+6 \gamma & =-1+\gamma(-1+6 \gamma) \\
-3+6 \gamma & =-1-\gamma+6 \gamma^{2} \\
6 \gamma^{2}-7 \gamma+2 & =0
\end{aligned}
$$

Example II, $\gamma=$ ?

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit,6) |

each field in the table is an n-tuple ( $s, a, s^{\prime}, r$ )


- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

$$
\begin{aligned}
q(C, \rightarrow) & =q(C, \leftarrow) \\
-3+6 \gamma & =-1+\gamma(-1+6 \gamma) \\
-3+6 \gamma & =-1-\gamma+6 \gamma^{2} \\
6 \gamma^{2}-7 \gamma+2 & =0 \\
\Rightarrow \gamma_{1}=2 / 3, \gamma_{2} & =1 / 2
\end{aligned}
$$

Example II, $\gamma=$ ?

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple ( $s, a, s^{\prime}, r$ )


- for $\gamma=1: \pi(B)=\leftarrow, \pi(C)=\leftarrow$
- Task: determine $\gamma$ which changes the policy computed for $\gamma=1$

$$
\begin{aligned}
& q(C, \rightarrow)=q(C, \leftarrow) \\
&-3+6 \gamma=-1+\gamma(-1+6 \gamma) \\
&-3+6 \gamma=-1-\gamma+6 \gamma^{2} \\
& 6 \gamma^{2}-7 \gamma+2=0 \\
& \Rightarrow \gamma_{1}=2 / 3, \gamma_{2}=1 / 2 \\
&\Rightarrow \pi(B)=\leftarrow, \pi(C)=\leftarrow ; \text { for } \gamma \in] 0,1 / 2[\cup] 2 / 3,1[ \\
&\pi(B)=\leftarrow, \pi(C)=\rightarrow ; \text { for } \gamma \in] 1 / 2,2 / 3[
\end{aligned}
$$

Example II

$$
\gamma=0
$$

## Example II, $\gamma=0$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

| A | B | C | D |
| :--- | :--- | :--- | :--- |

$\mathrm{S}=\{\mathrm{A}, B, \mathrm{C}, \mathrm{D}\}$
$A=\{\rightarrow, \leftarrow\}$
$r(\{B, C\}, \leftarrow)=-1, \quad r(\{A, D\})=6$,
$r(\{B, C\}, \rightarrow)=-3$
$p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2=$
$p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1$

For $\gamma=0$ :
Compute:
A: $q(B, \leftarrow)=6$
B: $q(B, \leftarrow)=5$
C: $q(B, \leftarrow)=-1$
D: $q(B, \leftarrow)=0$

## Example II, $\gamma=0$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

$$
\begin{aligned}
& \begin{array}{|c|c|c|c|}
\hline \mathrm{A} & \mathrm{~B} & \mathrm{C} & \mathrm{D} \\
\hline S=\{A, B, C, D\} \\
A=\{\rightarrow, \leftarrow\} \\
r(\{B, C\}, \leftarrow)=-1, \quad r(\{A, D\})=6, \\
r(\{B, C\}, \rightarrow)=-3 \\
p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\
p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1
\end{array}
\end{aligned}
$$

For $\gamma=0$ :
Compute:

A:

B:
$C: q(B, \leftarrow)=r(B, \leftarrow)+0 \cdot V(A)=-1$
D:

Example II, $\gamma=0$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  | $(A, \leftarrow$, exit, 6) |  |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

For $\gamma=0$ :

- $q(B, \leftarrow)=-1$


## Compute:

A: $q(B, \rightarrow)=6$
B: $q(B, \rightarrow)=-3$
C: $q(B, \rightarrow)=3$
D: $q(B, \rightarrow)=0$

Example II, $\gamma=0$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

$$
\begin{aligned}
& \begin{array}{l|c|c|c|}
\hline \mathrm{A} & \mathrm{~B} & \mathrm{C} & \mathrm{D} \\
\hline S=\{A, B, C, D\} \\
A=\{\rightarrow, \leftarrow\} \\
r(\{B, C\}, \leftarrow)=-1, r(\{A, D\})=6, \\
r(\{B, C\}, \rightarrow)=-3
\end{array} \\
& \begin{array}{l}
p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\
p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1
\end{array}
\end{aligned}
$$

$$
\text { For } \gamma=0
$$

$$
-q(B, \leftarrow)=-1
$$

## Compute:

A:
B: $q(B, \rightarrow)=r(B, \rightarrow)+0 \cdot V(C)=-3$
C:
D:

## Example II, $\gamma=0$

| Episode 1 Episode 2 Episode 3 Episode 4 <br> $(B, \rightarrow, C,-3)$ $(B, \leftarrow, A,-1)$ $(C, \rightarrow, D,-3)$ $(C, \leftarrow, B,-1)$ <br> $(C, \rightarrow, D,-3)$ $(A, \rightarrow$, exit, 6) $(D, \rightarrow$, exit, 6) $(B, \rightarrow, C,-3)$ <br> $(D, \leftarrow$, exit, 6)   $(C, \leftarrow, B,-1)$ <br>    $(B, \leftarrow, A,-1)$ <br>   $(A, \leftarrow$, exit, 6)  |  |  |  |
| :--- | :---: | :---: | :---: |
| each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$ |  |  |  |


| A | B | C | D |
| :--- | :--- | :--- | :--- |

$\mathrm{S}=\{A, B, C, D\}$
$A=\{\rightarrow, \leftarrow\}$
$r(\{B, C\}, \leftarrow)=-1, \quad r(\{A, D\})=6$,
$r(\{B, C\}, \rightarrow)=-3$
$p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2=$
$p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1$

For $\gamma=0$ :

- $q(B, \leftarrow)=-1$
- $q(B, \rightarrow)=-3$


## Example II, $\gamma=0$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  | $(A, \leftarrow$, exit, 6) |  |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

$$
\begin{aligned}
& \begin{array}{|c|c|c|c|}
\hline \mathrm{A} & \mathrm{~B} & \mathrm{C} & \mathrm{D} \\
\hline S=\{A, B, C, D\} \\
A=\{\rightarrow, \leftarrow\} \\
r(\{B, C\}, \leftarrow)=-1, r(\{A, D\})=6, \\
r(\{B, C\}, \rightarrow)=-3 \\
p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\
p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1
\end{array}
\end{aligned}
$$

For $\gamma=0$ :

- $q(B, \leftarrow)=-1$
- $q(B, \rightarrow)=-3$
$\Rightarrow \pi(B)=\leftarrow$

Example II, $\gamma=0$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

$$
\begin{aligned}
& \begin{array}{|c|c|c|c|}
\hline \mathrm{A} & \mathrm{~B} & \mathrm{C} & \mathrm{D} \\
\hline S=\{A, B, C, D\} \\
A=\{\rightarrow, \leftarrow\} \\
r(\{B, C\}, \leftarrow)=-1, r(\{A, D\})=6, \\
r(\{B, C\}, \rightarrow)=-3 \\
p(C \mid B, \rightarrow)=p(A \mid B, \leftarrow)=2 / 2= \\
p(D \mid C, \rightarrow)=p(B \mid C, \leftarrow)=1
\end{array}
\end{aligned}
$$

$$
\text { For } \gamma=0
$$

- $\pi(B)=\leftarrow$

Example II, $\gamma=0$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

For $\gamma=0$ :

- $\pi(B)=\leftarrow$


## Compute:

A: $q(C, \leftarrow)=-3$
B: $q(C, \leftarrow)=-1$
$C: q(C, \leftarrow)=6$
D: $q(C, \leftarrow)=3$

Example II, $\gamma=0$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$

For $\gamma=0$ :

- $\pi(B)=\leftarrow$


## Compute:

A:
B: $q(C, \leftarrow)=r(C, \leftarrow)+0 . V(B)=-1$
C:

D:

## Example II, $\gamma=0$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple ( $s, a, s^{\prime}, r$ )


For $\gamma=0$ :

- $\pi(B)=\leftarrow$
$-q(C, \leftarrow)=-1$
Compute:
A: $q(C, \rightarrow)=-3$
B: $q(C, \rightarrow)=-1$
C: $q(C, \rightarrow)=6$
D: $q(C, \rightarrow)=3$


## Example II, $\gamma=0$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple ( $s, a, s^{\prime}, r$ )


For $\gamma=0$ :

- $\pi(B)=\leftarrow$
$-q(C, \leftarrow)=-1$
Compute:
A: $q(C, \rightarrow)=r(C, \rightarrow)+0 \cdot V(D)=-3$
$B$ :

C:
D:

## Example II, $\gamma=0$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple ( $s, a, s^{\prime}, r$ )
 For $\gamma=0$ :

- $\pi(B)=\leftarrow$
$-q(C, \leftarrow)=-1$
$-q(C, \rightarrow)=-3$


## Example II, $\gamma=0$

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple ( $s, a, s^{\prime}, r$ )


For $\gamma=0$ :

- $\pi(B)=\leftarrow$
$-q(C, \leftarrow)=-1$
- $q(C, \rightarrow)=-3$
$\Rightarrow \pi(C)=\leftarrow$


# Example II 

summary

## Example II, summary

| Episode 1 | Episode 2 | Episode 3 | Episode 4 |
| :---: | :---: | :---: | :---: |
| $(B, \rightarrow, C,-3)$ | $(B, \leftarrow, A,-1)$ | $(C, \rightarrow, D,-3)$ | $(C, \leftarrow, B,-1)$ |
| $(C, \rightarrow, D,-3)$ | $(A, \rightarrow$, exit, 6) | $(D, \rightarrow$, exit, 6) | $(B, \rightarrow, C,-3)$ |
| $(D, \leftarrow$, exit, 6) |  |  | $(C, \leftarrow, B,-1)$ |
|  |  |  | $(B, \leftarrow, A,-1)$ |
|  |  |  | $(A, \leftarrow$, exit, 6) |

each field in the table is an n-tuple $\left(s, a, s^{\prime}, r\right)$


For $\gamma=1$ :

$$
\begin{aligned}
& \pi(B)=\leftarrow \\
& \pi(C)=\leftarrow
\end{aligned}
$$

For $\gamma \in] 2 / 3,1]: \quad \pi(B)=\leftarrow$ $\pi(C)=\leftarrow$

For $\gamma \in] 1 / 2,2 / 3[: \quad \pi(B)=\leftarrow$
$\pi(C)=\rightarrow$
For $\gamma \in[0,1 / 2[: \quad \pi(B)=\leftarrow$
$\pi(C)=\leftarrow$
For $\gamma=0: \quad \pi(B)=\leftarrow$ $\pi(C)=\leftarrow$

