

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 γ settings
- ▶ Study the boundary values for γ

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 γ settings
- ▶ Study the boundary values for γ

Example I

Example 1

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

each field in the table is an n-tuple (s, a, s', r)

Compute policy with

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

Example 1

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

each field in the table is an n-tuple (s, a, s', r)

Compute policy with

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

Example 1

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

each field in the table is an n-tuple (s, a, s', r)

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

World structure:

A	B	C	D
---	---	---	---

Example 1

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

each field in the table is an n-tuple (s, a, s', r)

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

World structure:

A	B	C	D
---	---	---	---

Example 1

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

each field in the table is an n-tuple (s, a, s', r)

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

World structure:

A	B	C	D
---	---	---	---

Example 1

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

each field in the table is an n-tuple (s, a, s', r)

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

World structure:

A	B	C	D
---	---	---	---

Example 1

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

each field in the table is an n-tuple (s, a, s', r)

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

World structure:

A	B	C	D
---	---	---	---

Example 1

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

each field in the table is an n-tuple (s, a, s', r)

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

World structure:

A	B	C	D
---	---	---	---

Example 1

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

each field in the table is an n-tuple (s, a, s', r)

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

World structure:

A	B	C	D
---	---	---	---

Example 1

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

each field in the table is an n-tuple (s, a, s', r)

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

World structure:

A	B	C	D
---	---	---	---

Example 1

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

each field in the table is an n-tuple (s, a, s', r)

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

World structure:

A	B	C	D
---	---	---	---

Example 1

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

each field in the table is an n-tuple (s, a, s', r)

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

World structure:

A	B	C	D
---	---	---	---

Example 1

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

each field in the table is an n-tuple (s, a, s', r)

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

World structure:

A	B	C	D
---	---	---	---

Example I

$$\gamma = 1$$

Example 1, $\gamma = 1$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

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 1, $\gamma = 1$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

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 1, $\gamma = 1$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

Compute:

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

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

$$C: q(B, \leftarrow) = 9$$

$$D: q(B, \leftarrow) = 6$$

Example 1, $\gamma = 1$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

Compute:

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

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

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

$$D: q(B, \leftarrow) = 6$$

Example 1, $\gamma = 1$

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

each field in the table is an n-tuple (s, a, s', r)

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

Compute:

(What can we assume about $\pi(C)$?)

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

$$B: q(B, \rightarrow) = 4$$

$$C: q(B, \rightarrow) = 10$$

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

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

Example I, $\gamma = 1$

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

each field in the table is an n-tuple (s, a, s', r)

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

Compute:

(What can we assume about $\pi(C)$?)

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

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

$$C: q(B, \rightarrow) = 10$$

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

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

Example I, $\gamma = 1$

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

each field in the table is an n-tuple (s, a, s', r)

- $q(B, \leftarrow) = 9$
- $q(B, \rightarrow) = 4$

Compute:

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

$$B: \pi(B) = \rightarrow$$

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

Example 1, $\gamma = 1$

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

each field in the table is an n-tuple (s, a, s', r)

- $q(B, \leftarrow) = 9$
- $q(B, \rightarrow) = 4$

Compute:

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

$$B: \pi(B) = \rightarrow$$

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

Example I, $\gamma = 1$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

► $\pi(B) = \leftarrow$

Compute:

A: $q(C, \rightarrow) = -1$

B: $q(C, \rightarrow) = 5$

C: $q(C, \rightarrow) = 9$

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

Example 1, $\gamma = 1$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

► $\pi(B) = \leftarrow$

Compute:

$$A: q(C, \rightarrow) = -1$$

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

$$C: q(C, \rightarrow) = 9$$

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

Example 1, $\gamma = 1$

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

each field in the table is an n-tuple (s, a, s', r)

► $\pi(B) = \leftarrow$

- $q(C, \rightarrow) = 5$

Compute:

(What can we assume about $\pi(B)$?)

A: $q(C, \leftarrow) = -1$

B: $q(C, \leftarrow) = 6$

C: $q(C, \leftarrow) = 10$

D: $q(C, \leftarrow) = 8$

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

Example 1, $\gamma = 1$

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

each field in the table is an n-tuple (s, a, s', r)

► $\pi(B) = \leftarrow$

- $q(C, \rightarrow) = 5$

Compute:

(What can we assume about $\pi(B)$?)

A: $q(C, \leftarrow) = -1$

B: $q(C, \leftarrow) = 6$

C: $q(C, \leftarrow) = 10$

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

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

Example 1, $\gamma = 1$

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

each field in the table is an n-tuple (s, a, s', r)

► $\pi(B) = \leftarrow$

- $q(C, \rightarrow) = 5$

- $q(C, \leftarrow) = 8$

Compute:

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

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

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

Example 1, $\gamma = 1$

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

each field in the table is an n-tuple (s, a, s', r)

► $\pi(B) = \leftarrow$

- $q(C, \rightarrow) = 5$

- $q(C, \leftarrow) = 8$

Compute:

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

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

A	B	C	D
---	---	---	---

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

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

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

$$p(C|B, \rightarrow) = p(A|B, \leftarrow) = 2/2 =$$

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

Example I, $\gamma = 1$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

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 1, $\gamma = 1$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

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

$$\gamma = ?$$

Example I, $\gamma = ?$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

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

► Task: determine γ 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
$(B, \rightarrow, C, -1)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -1)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -1)$	$(A, \rightarrow, \text{exit}, 10)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -1)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 10)$

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

- ▶ for $\gamma = 1$: $\pi(B) = \leftarrow, \pi(C) = \leftarrow$
- ▶ Task: determine γ 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
$(B, \rightarrow, C, -1)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -1)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -1)$	$(A, \rightarrow, \text{exit}, 10)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -1)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 10)$

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

- ▶ for $\gamma = 1$: $\pi(B) = \leftarrow, \pi(C) = \leftarrow$
- ▶ Task: determine γ 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
$(B, \rightarrow, C, -1)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -1)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -1)$	$(A, \rightarrow, \text{exit}, 10)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -1)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 10)$

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

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

How?

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

Example 1, $\gamma = ?$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

$p\{C|B, \rightarrow\} = p\{A|B, \leftarrow\} = 2/2 = 1$
 $p\{D|C, \rightarrow\} = p\{B|C, \leftarrow\} = 1$

- ▶ for $\gamma = 1$: $\pi(B) = \leftarrow, \pi(C) = \leftarrow$
- ▶ Task: determine γ 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 1, $\gamma = ?$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

- ▶ for $\gamma = 1$: $\pi(B) = \leftarrow, \pi(C) = \leftarrow$
- ▶ Task: determine γ 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 1, $\gamma = ?$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

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

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

Can the policy in state B be changed?

Compute:

$$A: V(B) = \sum_{s'} p(s'|B, a) \{V(s')\}, s' \in \{A, C\}$$

$$B: V(B) = \max_a (r(B) + \gamma \cdot V(s')), s' \in \{A, C\}$$

$$C: V(B) = \arg \max_a \sum_{s'} \gamma \cdot V(s'), s' \in \{A, C\}$$

$$D: V(B) = r(B) + \gamma V(D)$$

Example 1, $\gamma = ?$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

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

Can the policy in state B be changed?

Compute:

- A: $V(B) = \sum_{s'} p(s'|B, a) \{V(s')\}, s' \in \{A, C\}$
- B: $V(B) = \max_a (r(B) + \gamma \cdot V(s')), s' \in \{A, C\}$
- C: $V(B) = \arg \max_a \sum_{s'} \gamma \cdot V(s'), s' \in \{A, C\}$
- D: $V(B) = r(B) + \gamma V(D)$

Example 1, $\gamma = ?$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

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

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

Can the policy in state B be changed?

Compute:

$$A: V(B) = \sum_{s'} p(s'|B, a) \{V(s')\}, s' \in \{A, C\}$$

$$B: V(B) = \max_a (r(B) + \gamma \cdot V(s')), s' \in \{A, C\}$$

$$C: V(B) = \arg \max_a \sum_{s'} \gamma \cdot V(s'), s' \in \{A, C\}$$

$$D: V(B) = r(B) + \gamma V(D)$$

⇒ depends on $V(A), V(C)$

Example 1, $\gamma = ?$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

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

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

Can the policy in state B be changed?

Compute:

$$A: V(B) = \sum_{s'} p(s'|B, a) \{V(s')\}, s' \in \{A, C\}$$

$$B: V(B) = \max_a (r(B) + \gamma \cdot V(s')), s' \in \{A, C\}$$

$$C: V(B) = \arg \max_a \sum_{s'} \gamma \cdot V(s'), s' \in \{A, C\}$$

$$D: V(B) = r(B) + \gamma V(D)$$

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

Example I, $\gamma = ?$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

- ▶ for $\gamma = 1$: $\pi(B) = \leftarrow, \pi(C) = \leftarrow$
- ▶ Task: determine γ 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 1, $\gamma = ?$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

- ▶ for $\gamma = 1$: $\pi(B) = \leftarrow, \pi(C) = \leftarrow$
- ▶ Task: determine γ 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: $V(A) > V(C); V(A) = 10, V(C) < 10$

$\Rightarrow \pi(B) = \leftarrow$

$$V(B) = r(B) + \gamma \cdot V(A) = -1 + 10\gamma$$

Example 1, $\gamma = ?$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

- ▶ for $\gamma = 1$: $\pi(B) = \leftarrow, \pi(C) = \leftarrow$
- ▶ Task: determine γ 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: $V(A) > V(C); V(A) = 10, V(C) < 10$

$\Rightarrow \pi(B) = \leftarrow$

$$V(B) = r(B) + \gamma \cdot V(A) = -1 + 10\gamma$$

Example 1, $\gamma = ?$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

$$p(C|B, \rightarrow) = p(A|B, \leftarrow) = 2/2 =$$

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

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

► Task: determine γ 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 1, $\gamma = ?$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

- ▶ for $\gamma = 1$: $\pi(B) = \leftarrow, \pi(C) = \leftarrow$
- ▶ Task: determine γ 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 1, $\gamma = ?$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

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

► Task: determine γ 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 1, $\gamma = ?$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

- ▶ for $\gamma = 1$: $\pi(B) = \leftarrow, \pi(C) = \leftarrow$
- ▶ Task: determine γ 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 1, $\gamma = ?$

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

each field in the table is an n-tuple (s, a, s', r)

- ▶ for $\gamma = 1$: $\pi(B) = \leftarrow, \pi(C) = \leftarrow$
- ▶ Task: determine γ 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: $Q(C, \rightarrow) = r(C) + V(B)$

A	B	C	D
---	---	---	---

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

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

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

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

Example 1, $\gamma = ?$

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

each field in the table is an n-tuple (s, a, s', r)

- ▶ for $\gamma = 1$: $\pi(B) = \leftarrow, \pi(C) = \leftarrow$
- ▶ Task: determine γ 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:

$$A: Q(C, \rightarrow) = r(C) + V(D)$$

$$B: Q(C, \rightarrow) = r(C) + \gamma \cdot V(B)$$

$$C: Q(C, \rightarrow) = r(C) + \gamma \cdot V(D) = -1 + 6\gamma$$

$$D: Q(C, \rightarrow) = r(C) + V(B)$$

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

Example 1, $\gamma = ?$

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

each field in the table is an n-tuple (s, a, s', r)

- ▶ for $\gamma = 1$: $\pi(B) = \leftarrow, \pi(C) = \leftarrow$
- ▶ Task: determine γ 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:

$$A: Q(C, \leftarrow) = r(C) + V(D)$$

$$B: Q(C, \leftarrow) = r(C) + \gamma \cdot V(B)$$

$$C: Q(C, \leftarrow) = r(C) + \gamma \cdot V(D)$$

$$D: Q(C, \leftarrow) = r(C) + V(B)$$

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

Example 1, $\gamma = ?$

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

each field in the table is an n-tuple (s, a, s', r)

- ▶ for $\gamma = 1$: $\pi(B) = \leftarrow, \pi(C) = \leftarrow$
- ▶ Task: determine γ 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:

$$A: Q(C, \leftarrow) = r(C) + V(D)$$

$$B: Q(C, \leftarrow) = r(C) + \gamma \cdot V(B) = -1 + \gamma(-1 + 10\gamma)$$

$$C: Q(C, \leftarrow) = r(C) + \gamma \cdot V(D)$$

$$D: Q(C, \leftarrow) = r(C) + V(B)$$

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

Example 1, $\gamma = ?$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

- ▶ for $\gamma = 1$: $\pi(B) = \leftarrow, \pi(C) = \leftarrow$
- ▶ Task: determine γ 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, $\gamma = ?$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

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

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

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

$$-1 + 6\gamma > -1 + \gamma(-1 + 10\gamma)$$

$$-1 + 6\gamma > -1 - \gamma + 10\gamma^2$$

$$7\gamma - 10\gamma^2 > 0$$

$$\Rightarrow \gamma_1 = 0, \gamma_2 = 0.7$$

$$\Rightarrow \pi(B) = \leftarrow, \pi(C) = \leftarrow; \text{ for } \gamma \in]0.7, 1]$$

$$\pi(B) = \leftarrow, \pi(C) = \rightarrow; \text{ for } \gamma \in]0, 0.7[$$

Example I, $\gamma = ?$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

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

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

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

$$-1 + 6\gamma > -1 + \gamma(-1 + 10\gamma)$$

$$-1 + 6\gamma > -1 - \gamma + 10\gamma^2$$

$$7\gamma - 10\gamma^2 > 0$$

$$\Rightarrow \gamma_1 = 0, \gamma_2 = 0.7$$

$$\Rightarrow \pi(B) = \leftarrow, \pi(C) = \leftarrow; \text{ for } \gamma \in]0.7, 1]$$

$$\pi(B) = \leftarrow, \pi(C) = \rightarrow; \text{ for } \gamma \in]0, 0.7[$$

Example I, $\gamma = ?$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

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

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

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

$$-1 + 6\gamma > -1 + \gamma(-1 + 10\gamma)$$

$$-1 + 6\gamma > -1 - \gamma + 10\gamma^2$$

$$7\gamma - 10\gamma^2 > 0$$

$$\Rightarrow \gamma_1 = 0, \gamma_2 = 0.7$$

$$\Rightarrow \pi(B) = \leftarrow, \pi(C) = \leftarrow; \text{ for } \gamma \in]0.7, 1]$$

$$\pi(B) = \leftarrow, \pi(C) = \rightarrow; \text{ for } \gamma \in]0, 0.7[$$

Example I, $\gamma = ?$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

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

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

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

$$-1 + 6\gamma > -1 + \gamma(-1 + 10\gamma)$$

$$-1 + 6\gamma > -1 - \gamma + 10\gamma^2$$

$$7\gamma - 10\gamma^2 > 0$$

$$\Rightarrow \gamma_1 = 0, \gamma_2 = 0.7$$

$$\Rightarrow \pi(B) = \leftarrow, \pi(C) = \leftarrow; \text{ for } \gamma \in]0.7, 1]$$

$$\pi(B) = \leftarrow, \pi(C) = \rightarrow; \text{ for } \gamma \in]0, 0.7[$$

Example 1, $\gamma = ?$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

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

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

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

$$-1 + 6\gamma > -1 + \gamma(-1 + 10\gamma)$$

$$-1 + 6\gamma > -1 - \gamma + 10\gamma^2$$

$$7\gamma - 10\gamma^2 > 0$$

$$\Rightarrow \gamma_1 = 0, \gamma_2 = 0.7$$

$\Rightarrow \pi(B) = \leftarrow, \pi(C) = \leftarrow; \text{ for } \gamma \in]0.7, 1]$

$\pi(B) = \leftarrow, \pi(C) = \rightarrow; \text{ for } \gamma \in]0, 0.7[$

Example 1, $\gamma = ?$

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

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

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

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

$$-1 + 6\gamma > -1 + \gamma(-1 + 10\gamma)$$

$$-1 + 6\gamma > -1 - \gamma + 10\gamma^2$$

$$7\gamma - 10\gamma^2 > 0$$

$$\Rightarrow \gamma_1 = 0, \gamma_2 = 0.7$$

$\Rightarrow \pi(B) = \leftarrow, \pi(C) = \leftarrow; \text{ for } \gamma \in]0.7, 1]$

$\pi(B) = \leftarrow, \pi(C) = \rightarrow; \text{ for } \gamma \in]0, 0.7[$

Example I

$$\gamma = 0$$

Example I, $\gamma = 0$

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

each field in the table is an n-tuple (s, a, s', r)

For $\gamma = 0$:

Compute:

$$A: V(B) = 9$$

$$B: V(B) = 5$$

$$C: V(B) = -1$$

$$D: V(B) = 0$$

A	B	C	D
---	---	---	---

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

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

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

$$p(C|B, \rightarrow) = p(A|B, \leftarrow) = 2/2 =$$

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

Example I, $\gamma = 0$

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

each field in the table is an n-tuple (s, a, s', r)

For $\gamma = 0$:

Compute:

$$A: V(B) = 9$$

$$B: V(B) = 5$$

$$C: V(B) = r(B) + \gamma \max_a V(s') = -1 + 0 = -1$$

$$D: V(B) = 0$$

A	B	C	D
---	---	---	---

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2}{2} = 1$$

Example I, $\gamma = 0$

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

each field in the table is an n-tuple (s, a, s', r)

For $\gamma = 0$:

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

Compute:

$$A: V(C) = 9$$

$$B: V(C) = 5$$

$$C: V(C) = -1$$

$$D: V(C) = 0$$

A	B	C	D
---	---	---	---

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

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

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

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

Example 1, $\gamma = 0$

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

each field in the table is an n-tuple (s, a, s', r)

For $\gamma = 0$:

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

Compute:

$$A: V(C) = 9$$

$$B: V(C) = 5$$

$$C: V(C) = r(C) + \gamma \max_a V(s') = -1 + 0 = -1$$

$$D: V(C) = 0$$

A	B	C	D
---	---	---	---

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

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

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

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

Example I, $\gamma = 0$

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

each field in the table is an n-tuple (s, a, s', r)

For $\gamma = 0$:

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

$$\Rightarrow \pi(B) = \{\leftarrow, \rightarrow\}$$

$$\pi(C) = \{\leftarrow, \rightarrow\}$$

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

Example I, $\gamma = 0$

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

each field in the table is an n-tuple (s, a, s', r)

For $\gamma = 0$:

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

$$\Rightarrow \pi(B) = \{\leftarrow, \rightarrow\}$$

$$\pi(C) = \{\leftarrow, \rightarrow\}$$

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

Example I

summary

Example I, summary

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

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

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

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

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

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

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

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

$$\text{For } \gamma = 0: \quad \pi(B) = \{\leftarrow, \rightarrow\}$$

$$\pi(C) = \{\leftarrow, \rightarrow\}$$

Example II

Example II

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

Compute policy with

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

Example II

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

Compute policy with

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

Example II

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

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

World structure:

A	B	C	D
---	---	---	---

Example II

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

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

World structure:

A	B	C	D
---	---	---	---

Example II

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

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

World structure:

A	B	C	D
---	---	---	---

Example II

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

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

World structure:

A	B	C	D
---	---	---	---

Example II

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

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

World structure:

A	B	C	D
---	---	---	---

Example II

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

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

World structure:

A	B	C	D
---	---	---	---

Example II

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

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

World structure:

A	B	C	D
---	---	---	---

Example II

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

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

World structure:

A	B	C	D
---	---	---	---

Example II

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

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

World structure:

A	B	C	D
---	---	---	---

Example II

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

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

World structure:

A	B	C	D
---	---	---	---

Example II

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

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

World structure:

A	B	C	D
---	---	---	---

Example II

$$\gamma = 1$$

Example II, $\gamma = 1$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2}{2} = 1$$

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
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2/2}{1} = 1$$

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
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

$$p(C|B, \rightarrow) = p(A|B, \leftarrow) = 2/2 =$$

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

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
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

$$p(C|B, \rightarrow) = p(A|B, \leftarrow) = 2/2 =$$

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

Compute:

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

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

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

$$D: q(B, \leftarrow) = 6$$

Example II, $\gamma = 1$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

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

Compute:

(What can we assume about $\pi(C)$?)

$$\text{A: } q(B, \rightarrow) = -1$$

$$\text{B: } q(B, \rightarrow) = 4$$

$$\text{C: } q(B, \rightarrow) = 0$$

$$\text{D: } q(B, \rightarrow) = -3$$

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2}{2} = 1$$

Example II, $\gamma = 1$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

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

Compute:

(What can we assume about $\pi(C)$?)

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

$$B: q(B, \rightarrow) = 4$$

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

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

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2}{2} = 1$$

Example II, $\gamma = 1$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

- $q(B, \leftarrow) = 5$
- $q(B, \rightarrow) = 0$

Compute:

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

B: $\pi(B) = \rightarrow$

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2}{2} = 1$$

Example II, $\gamma = 1$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

- $q(B, \leftarrow) = 5$
- $q(B, \rightarrow) = 0$

Compute:

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

B: $\pi(B) = \rightarrow$

A	B	C	D
---	---	---	---

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

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

Example II, $\gamma = 1$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

► $\pi(B) = \leftarrow$

Compute:

A: $q(C, \rightarrow) = -1$

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

C: $q(C, \rightarrow) = 3$

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

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2}{2} = 1$$

Example II, $\gamma = 1$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

► $\pi(B) = \leftarrow$

Compute:

A: $q(C, \rightarrow) = -1$

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

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

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

A	B	C	D
---	---	---	---

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

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

Example II, $\gamma = 1$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

► $\pi(B) = \leftarrow$

- $q(C, \rightarrow) = 3$

Compute:

(What can we assume about $\pi(B)$?)

A: $q(C, \leftarrow) = -1$

B: $q(C, \leftarrow) = 6$

C: $q(C, \leftarrow) = -3$

D: $q(C, \leftarrow) = 4$

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2}{2} = 1$$

Example II, $\gamma = 1$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

► $\pi(B) = \leftarrow$

- $q(C, \rightarrow) = 3$

Compute:

(What can we assume about $\pi(B)$?)

A: $q(C, \leftarrow) = -1$

B: $q(C, \leftarrow) = 6$

C: $q(C, \leftarrow) = -3$

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

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2/2}{1} = 1$$

Example II, $\gamma = 1$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

► $\pi(B) = \leftarrow$

- $q(C, \rightarrow) = 3$

- $q(C, \leftarrow) = 4$

Compute:

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

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

A	B	C	D
---	---	---	---

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

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

Example II, $\gamma = 1$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

- ▶ $\pi(B) = \leftarrow$
- $q(C, \rightarrow) = 3$
- $q(C, \leftarrow) = 4$

Compute:

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

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

A	B	C	D
---	---	---	---

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

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

Example II, $\gamma = 1$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2}{2} = 1$$

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
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2}{2} = 1$$

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

$$\gamma = ?$$

Example II, $\gamma = ?$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

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

Can the policy in state B be changed?

A: Yes

B: No

Let's find out :-)

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2}{2} = 1$$

Example II, $\gamma = ?$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2}{2} = 1$$

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

► Task: determine γ 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
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

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

Can the policy in state B be changed?

$$\begin{aligned}
 V(B) &= \max_a (r(B, a) + \gamma \cdot V(s')), s' \in \{A, C\} \\
 &= \max \left\{ \begin{array}{l} (\rightarrow) \quad r(B, \rightarrow) + \gamma \cdot V(C) \\ (\leftarrow) \quad r(B, \leftarrow) + \gamma \cdot V(A) \end{array} \right\} \\
 &= r(B, \leftarrow) + \gamma \cdot V(A) \text{ since } V(A) > V(C) \ \& \ r(B, \leftarrow) > r(B, \rightarrow)
 \end{aligned}$$

$$\Rightarrow \pi(B) = \leftarrow$$

$$V(B) = -1 + 6\gamma$$

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2}{2} = 1$$

Example II, $\gamma = ?$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

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

Can the policy in state B be changed?

$$V(B) = \max_a (r(B, a) + \gamma \cdot V(s')), s' \in \{A, C\}$$

$$= \max \left\{ \begin{array}{l} (\rightarrow) \quad r(B, \rightarrow) + \gamma \cdot V(C) \\ (\leftarrow) \quad r(B, \leftarrow) + \gamma \cdot V(A) \end{array} \right\}$$

$$= r(B, \leftarrow) + \gamma \cdot V(A) \text{ since } V(A) > V(C) \ \& \ r(B, \leftarrow) > r(B, \rightarrow)$$

$$\Rightarrow \pi(B) = \leftarrow$$

$$V(B) = -1 + 6\gamma$$

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2}{2} = 1$$

Example II, $\gamma = ?$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

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

Can the policy in state B be changed?

$$V(B) = \max_a (r(B, a) + \gamma \cdot V(s')), s' \in \{A, C\}$$

$$= \max \left\{ \begin{array}{l} (\rightarrow) \quad r(B, \rightarrow) + \gamma \cdot V(C) \\ (\leftarrow) \quad r(B, \leftarrow) + \gamma \cdot V(A) \end{array} \right\}$$

$$= r(B, \leftarrow) + \gamma \cdot V(A) \text{ since } V(A) > V(C) \ \& \ r(B, \leftarrow) > r(B, \rightarrow)$$

$$\Rightarrow \pi(B) = \leftarrow$$

$$V(B) = -1 + 6\gamma$$

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2}{2} = 1$$

Example II, $\gamma = ?$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

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

Can the policy in state B be changed?

$$\begin{aligned}
 V(B) &= \max_a (r(B, a) + \gamma \cdot V(s')), s' \in \{A, C\} \\
 &= \max \left\{ \begin{array}{l} (\rightarrow) \quad r(B, \rightarrow) + \gamma \cdot V(C) \\ (\leftarrow) \quad r(B, \leftarrow) + \gamma \cdot V(A) \end{array} \right\} \\
 &= r(B, \leftarrow) + \gamma \cdot V(A) \text{ since } V(A) > V(C) \ \& \ r(B, \leftarrow) > r(B, \rightarrow)
 \end{aligned}$$

$$\Rightarrow \pi(B) = \leftarrow$$

$$V(B) = -1 + 6\gamma$$

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2}{2} = 1$$

Example II, $\gamma = ?$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

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

Can the policy in state B be changed?

$$\begin{aligned}
 V(B) &= \max_a (r(B, a) + \gamma \cdot V(s')), s' \in \{A, C\} \\
 &= \max \left\{ \begin{array}{l} (\rightarrow) \quad r(B, \rightarrow) + \gamma \cdot V(C) \\ (\leftarrow) \quad r(B, \leftarrow) + \gamma \cdot V(A) \end{array} \right\} \\
 &= r(B, \leftarrow) + \gamma \cdot V(A) \text{ since } V(A) > V(C) \ \& \ r(B, \leftarrow) > r(B, \rightarrow)
 \end{aligned}$$

$$\Rightarrow \pi(B) = \leftarrow$$

$$V(B) = -1 + 6\gamma$$

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2/2}{1} = 1$$

Example II, $\gamma = ?$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

- ▶ for $\gamma = 1$: $\pi(B) = \leftarrow, \pi(C) = \leftarrow$
- ▶ Task: determine γ 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:

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

$$B: Q(C, \rightarrow) = r(C, \rightarrow) + \gamma \cdot V(B)$$

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

$$D: Q(C, \rightarrow) = r(C, \leftarrow) + V(B)$$

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2/2}{1} = 1$$

Example II, $\gamma = ?$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

- ▶ for $\gamma = 1$: $\pi(B) = \leftarrow, \pi(C) = \leftarrow$
- ▶ Task: determine γ 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:

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

$$B: Q(C, \rightarrow) = r(C, \rightarrow) + \gamma \cdot V(B)$$

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

$$D: Q(C, \rightarrow) = r(C, \leftarrow) + V(B)$$

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2/2}{1} = 1$$

Example II, $\gamma = ?$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

- ▶ for $\gamma = 1$: $\pi(B) = \leftarrow, \pi(C) = \leftarrow$
- ▶ Task: determine γ 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:

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

$$B: Q(C, \rightarrow) = r(C, \rightarrow) + \gamma \cdot V(B)$$

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

$$D: Q(C, \rightarrow) = r(C, \leftarrow) + V(B)$$

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2/2}{1} =$$

Example II, $\gamma = ?$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

- ▶ for $\gamma = 1$: $\pi(B) = \leftarrow, \pi(C) = \leftarrow$
- ▶ Task: determine γ 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:

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

$$B: Q(C, \rightarrow) = r(C, \rightarrow) + \gamma \cdot V(B)$$

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

$$D: Q(C, \rightarrow) = r(C, \leftarrow) + V(B)$$

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2}{2} = 1$$

Example II, $\gamma = ?$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

- ▶ for $\gamma = 1$: $\pi(B) = \leftarrow, \pi(C) = \leftarrow$
- ▶ Task: determine γ 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:

$$A: Q(C, \leftarrow) = r(C) + V(D)$$

$$B: Q(C, \leftarrow) = r(C, \leftarrow) + \gamma \cdot V(B)$$

$$C: Q(C, \leftarrow) = r(C, \leftarrow) + \gamma \cdot V(D)$$

$$D: Q(C, \leftarrow) = r(C, \rightarrow) + V(B)$$

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2}{2} = 1$$

Example II, $\gamma = ?$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

- ▶ for $\gamma = 1$: $\pi(B) = \leftarrow, \pi(C) = \leftarrow$
- ▶ Task: determine γ 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:

$$A: Q(C, \leftarrow) = r(C) + V(D)$$

$$B: Q(C, \leftarrow) = r(C, \leftarrow) + \gamma \cdot V(B) = -1 + \gamma(-1 + 6\gamma)$$

$$C: Q(C, \leftarrow) = r(C, \leftarrow) + \gamma \cdot V(D)$$

$$D: Q(C, \leftarrow) = r(C, \rightarrow) + V(B)$$

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2}{2} = 1$$

Example II, $\gamma = ?$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

- ▶ for $\gamma = 1$: $\pi(B) = \leftarrow, \pi(C) = \leftarrow$
- ▶ Task: determine γ 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:

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

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2/2}{1} = 1$$

Example II, $\gamma = ?$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

- ▶ for $\gamma = 1$: $\pi(B) = \leftarrow, \pi(C) = \leftarrow$
- ▶ Task: determine γ 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:

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

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2/2}{1} = 1$$

Example II, $\gamma = ?$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

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

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2}{2} = 1$$

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

$$-3 + 6\gamma = -1 + \gamma(-1 + 6\gamma)$$

$$-3 + 6\gamma = -1 - \gamma + 6\gamma^2$$

$$6\gamma^2 - 7\gamma + 2 = 0$$

$$\Rightarrow \gamma_1 = 2/3, \gamma_2 = 1/2$$

$$\Rightarrow \pi(B) = \leftarrow, \pi(C) = \leftarrow; \text{ for } \gamma \in]0, 1/2[\cup]2/3, 1[$$

$$\pi(B) = \leftarrow, \pi(C) = \rightarrow; \text{ for } \gamma \in]1/2, 2/3[$$

Example II, $\gamma = ?$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

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

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

$$-3 + 6\gamma = -1 + \gamma(-1 + 6\gamma)$$

$$-3 + 6\gamma = -1 - \gamma + 6\gamma^2$$

$$6\gamma^2 - 7\gamma + 2 = 0$$

$$\Rightarrow \gamma_1 = 2/3, \gamma_2 = 1/2$$

$$\Rightarrow \pi(B) = \leftarrow, \pi(C) = \leftarrow; \text{ for } \gamma \in]0, 1/2[\cup]2/3, 1[$$

$$\pi(B) = \leftarrow, \pi(C) = \rightarrow; \text{ for } \gamma \in]1/2, 2/3[$$

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2}{2} = 1$$

Example II, $\gamma = ?$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

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

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

$$\begin{aligned}
 6\gamma^2 - 7\gamma + 2 &= 0 \\
 \Rightarrow \gamma_1 = 2/3, \gamma_2 = 1/2
 \end{aligned}$$

$$\Rightarrow \pi(B) = \leftarrow, \pi(C) = \leftarrow; \text{ for } \gamma \in]0, 1/2[\cup]2/3, 1[$$

$$\pi(B) = \leftarrow, \pi(C) = \rightarrow; \text{ for } \gamma \in]1/2, 2/3[$$

A	B	C	D
---	---	---	---

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

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

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

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

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

Example II, $\gamma = ?$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

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

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

$$\Rightarrow \gamma_1 = 2/3, \gamma_2 = 1/2$$

$$\Rightarrow \pi(B) = \leftarrow, \pi(C) = \leftarrow; \text{ for } \gamma \in]0, 1/2[\cup]2/3, 1[$$

$$\pi(B) = \leftarrow, \pi(C) = \rightarrow; \text{ for } \gamma \in]1/2, 2/3[$$

A	B	C	D
---	---	---	---

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

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

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

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

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

Example II, $\gamma = ?$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

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

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

$$6\gamma^2 - 7\gamma + 2 = 0$$

$$\Rightarrow \gamma_1 = 2/3, \gamma_2 = 1/2$$

$\Rightarrow \pi(B) = \leftarrow, \pi(C) = \leftarrow$; for $\gamma \in]0, 1/2[\cup]2/3, 1[$

$\pi(B) = \leftarrow, \pi(C) = \rightarrow$; for $\gamma \in]1/2, 2/3[$

A	B	C	D
---	---	---	---

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

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

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

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

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

Example II, $\gamma = ?$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

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

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

$$6\gamma^2 - 7\gamma + 2 = 0$$

$$\Rightarrow \gamma_1 = 2/3, \gamma_2 = 1/2$$

$$\Rightarrow \pi(B) = \leftarrow, \pi(C) = \leftarrow; \text{ for } \gamma \in]0, 1/2[\cup]2/3, 1[$$

$$\pi(B) = \leftarrow, \pi(C) = \rightarrow; \text{ for } \gamma \in]1/2, 2/3[$$

A	B	C	D
---	---	---	---

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

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

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

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

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

Example II

$$\gamma = 0$$

Example II, $\gamma = 0$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

For $\gamma = 0$:

Compute:

$$A: Q(B, \leftarrow) = 6$$

$$B: Q(B, \leftarrow) = 5$$

$$C: Q(B, \leftarrow) = -1$$

$$D: Q(B, \leftarrow) = 0$$

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2}{2} = 1$$

Example II, $\gamma = 0$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

For $\gamma = 0$:

Compute:

$$A: Q(B, \leftarrow) = 6$$

$$B: Q(B, \leftarrow) = 5$$

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

$$D: Q(B, \leftarrow) = 0$$

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2}{2} = 1$$

Example II, $\gamma = 0$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

For $\gamma = 0$:

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

Compute:

$$A: Q(B, \rightarrow) = 6$$

$$B: Q(B, \rightarrow) = -3$$

$$C: Q(B, \rightarrow) = 3$$

$$D: Q(B, \rightarrow) = 0$$

A	B	C	D
---	---	---	---

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

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

Example II, $\gamma = 0$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

For $\gamma = 0$:

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

Compute:

$$A: Q(B, \rightarrow) = 6$$

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

$$C: Q(B, \rightarrow) = 3$$

$$D: Q(B, \rightarrow) = 0$$

A	B	C	D
---	---	---	---

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

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

Example II, $\gamma = 0$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

For $\gamma = 0$:

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

$\Rightarrow \pi(B) = \leftarrow$

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2}{2} = 1$$

Example II, $\gamma = 0$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

For $\gamma = 0$:

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

$\Rightarrow \pi(B) = \leftarrow$

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2}{2} = 1$$

Example II, $\gamma = 0$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

For $\gamma = 0$:

► $\pi(B) = \leftarrow$

Compute:

A: $Q(C, \leftarrow) = -3$

B: $Q(C, \leftarrow) = -1$

C: $Q(C, \leftarrow) = 6$

D: $Q(C, \leftarrow) = 3$

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2}{2} = 1$$

Example II, $\gamma = 0$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

For $\gamma = 0$:

► $\pi(B) = \leftarrow$

Compute:

A: $Q(C, \leftarrow) = -3$

B: $Q(C, \leftarrow) = -1$

C: $Q(C, \leftarrow) = 6$

D: $Q(C, \leftarrow) = 3$

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2/2}{1} = 1$$

Example II, $\gamma = 0$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

For $\gamma = 0$:

► $\pi(B) = \leftarrow$

Compute:

A: $Q(C, \leftarrow) = -3$

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

C: $Q(C, \leftarrow) = 6$

D: $Q(C, \leftarrow) = 3$

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2}{2} = 1$$

Example II, $\gamma = 0$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', 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: $Q(C, \rightarrow) = 3$

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2/2}{1} = 1$$

Example II, $\gamma = 0$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

For $\gamma = 0$:

- ▶ $\pi(B) = \leftarrow$
- $Q(C, \leftarrow) = -1$

Compute:

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

$$B: Q(C, \rightarrow) = -1$$

$$C: Q(C, \rightarrow) = 6$$

$$D: Q(C, \rightarrow) = 3$$

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2/2}{1} = 1$$

Example II, $\gamma = 0$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

For $\gamma = 0$:

► $\pi(B) = \leftarrow$

- $Q(C, \leftarrow) = -1$

- $Q(C, \rightarrow) = -3$

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

A	B	C	D
---	---	---	---

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

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

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

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

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

$$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$$

Example II, $\gamma = 0$

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

For $\gamma = 0$:

► $\pi(B) = \leftarrow$

- $Q(C, \leftarrow) = -1$

- $Q(C, \rightarrow) = -3$

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

A	B	C	D
---	---	---	---

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

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

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

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

$p(C|B, \rightarrow) = p(A|B, \leftarrow) = 2/2 =$

$p(D|C, \rightarrow) = p(B|C, \leftarrow) = 1$

Example II

summary

Example II, summary

Episode 1	Episode 2	Episode 3	Episode 4
$(B, \rightarrow, C, -3)$	$(B, \leftarrow, A, -1)$	$(C, \rightarrow, D, -3)$	$(C, \leftarrow, B, -1)$
$(C, \rightarrow, D, -3)$	$(A, \rightarrow, \text{exit}, 6)$	$(D, \rightarrow, \text{exit}, 6)$	$(B, \rightarrow, C, -3)$
$(D, \leftarrow, \text{exit}, 6)$			$(C, \leftarrow, B, -1)$
			$(B, \leftarrow, A, -1)$
			$(A, \leftarrow, \text{exit}, 6)$

each field in the table is an n-tuple (s, a, s', r)

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

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

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

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

$$\text{For } \gamma \in]1/2, 2/3[: \quad \pi(B) = \leftarrow$$

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

$$\text{For } \gamma \in [0, 1/2[: \quad \pi(B) = \leftarrow$$

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

$$\text{For } \gamma = 0: \quad \pi(B) = \leftarrow$$

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

A	B	C	D
---	---	---	---

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

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

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

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

$$\frac{p(C|B, \rightarrow)}{p(D|C, \rightarrow)} = \frac{p(A|B, \leftarrow)}{p(B|C, \leftarrow)} = \frac{2}{2} = 1$$