# Example

- Two states: A, B
- Two actions: Up, Down
- Discount factor: $\gamma = 0.5$
- Learning rate: $\alpha = 0.5$
- Q(A, Down) = ?
- Q(B, Up) = ?

| $t$ | $s_t$ | $a_t$ | $s_{t+1}$ | $r_t$ |
|---|---|---|---|---|
| 0 | A | Down | B | 2 |
| 1 | B | Down | B | -4 |
| 2 | B | Up | B | 0 |
| 3 | B | Up | A | 3 |
| 4 | A | Up | A | -1 |

$$Q(s,a) \leftarrow (1-\alpha)Q(s,a) + (\alpha)\left[r + \gamma \max_{a'} Q(s',a')\right]$$