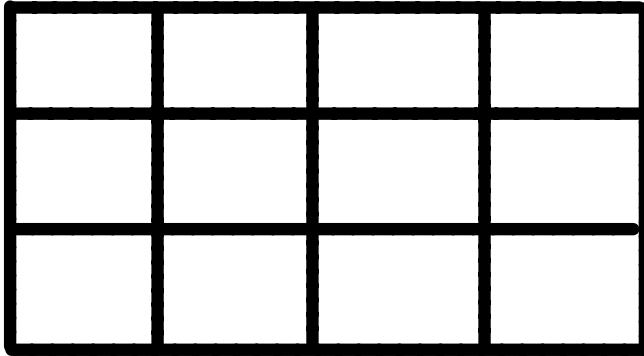


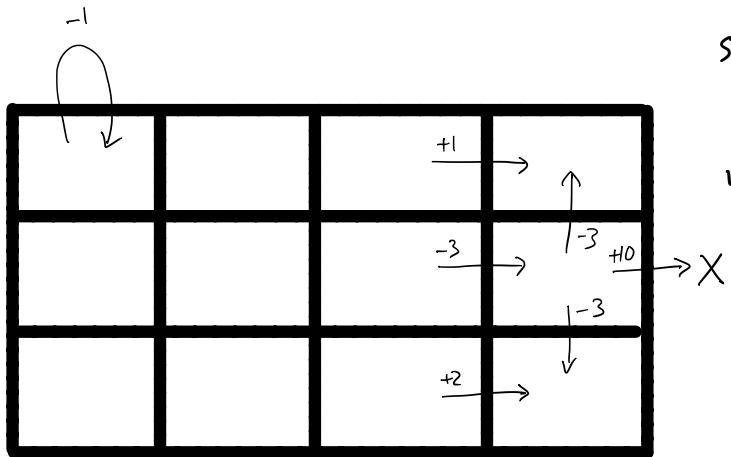
Stand



Clap

Wave

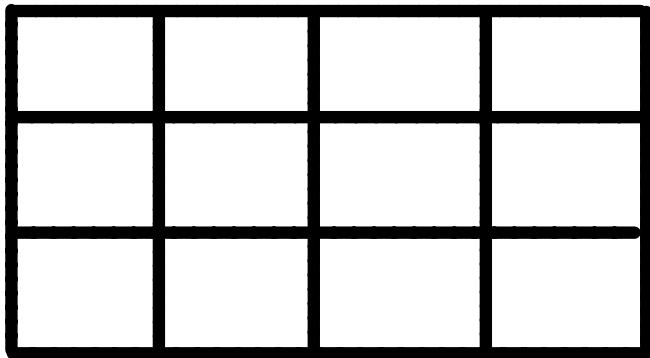
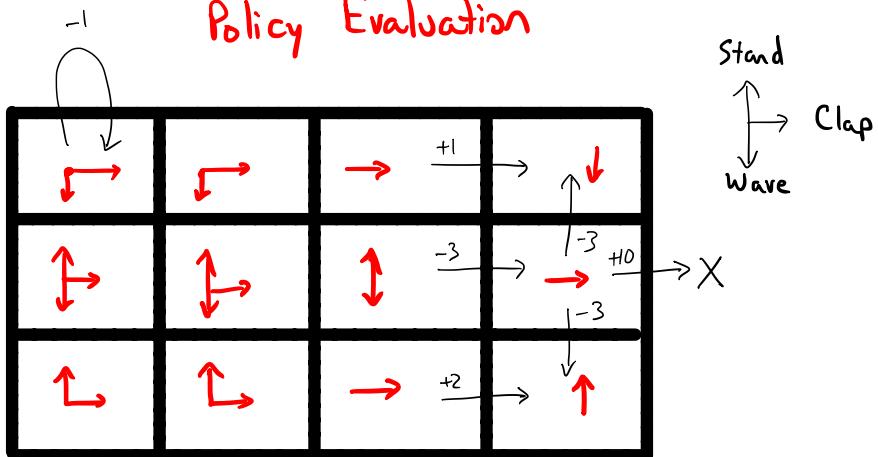
(RL: no model)



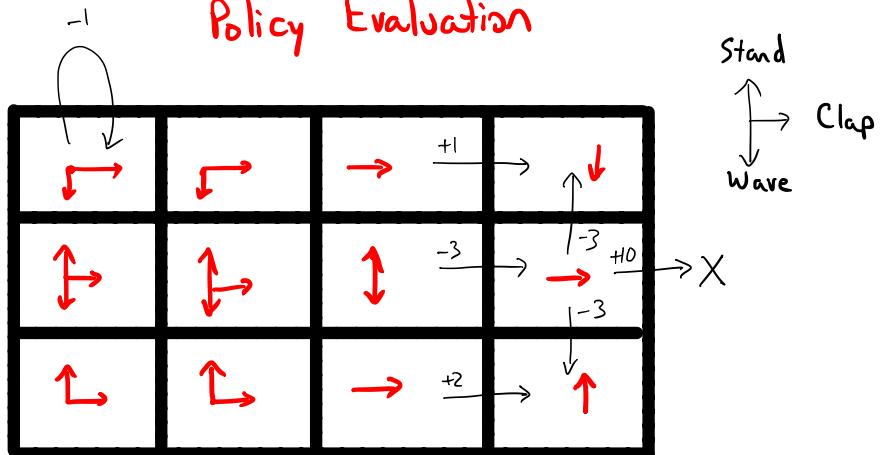
Stand  
Clap  
Wave

(DP: Known model)

## Policy Evaluation



## Policy Evaluation

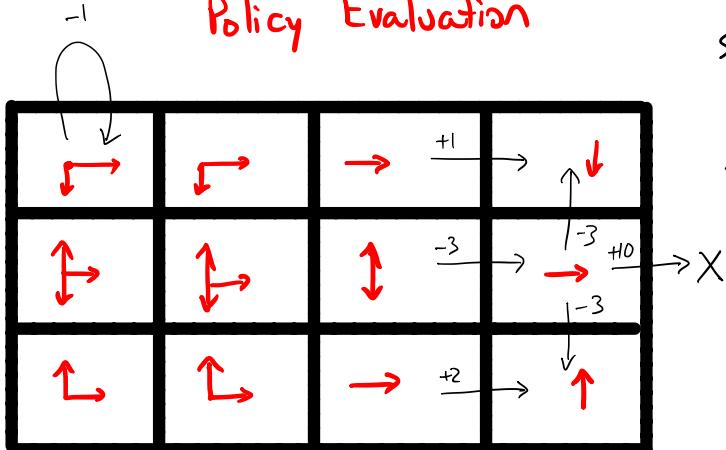


0	0	0	0
0	0	0	0
0	0	0	0

As  $v_b(s) = 0$   
 $v_{k+1}(s) =$        $(E_F, 4, 5)$

↑  
required

## Policy Evaluation

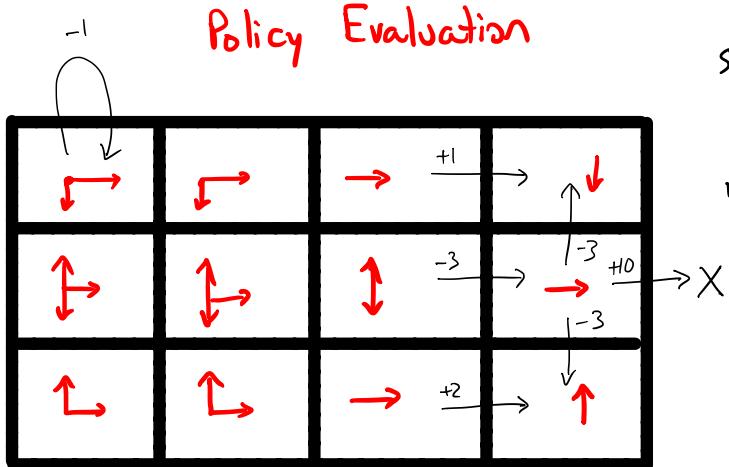


Stand  
Clap  
Wave

0	0	0	0
0	0	0	0
0	0	0	0

As  $v_0(s) = 0$   
 $v_{k+1}^{(s)} = \sum_a \pi(a|s) \times ?$

## Policy Evaluation



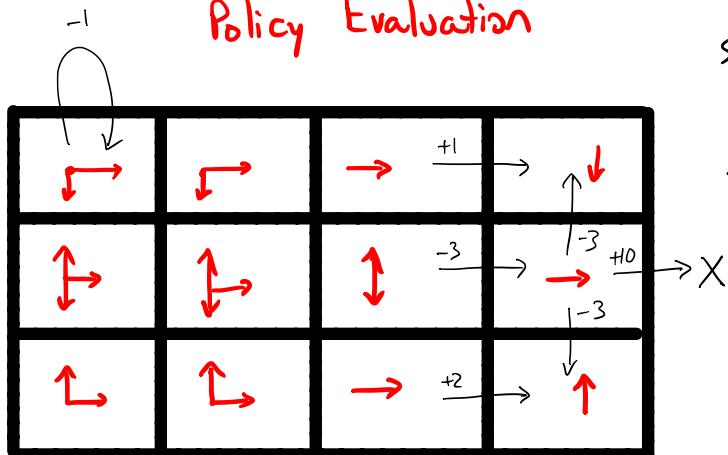
0	0	0	0
0	0	0	0
0	0	0	0

Stand  
Clap  
Wave

$$\forall s \quad v_\pi(s) = 0$$

$$v_{k+1}(s) = \sum_a \pi(a|s) \sum_{s',r} p(s',r|s,a) \times ?$$

## Policy Evaluation



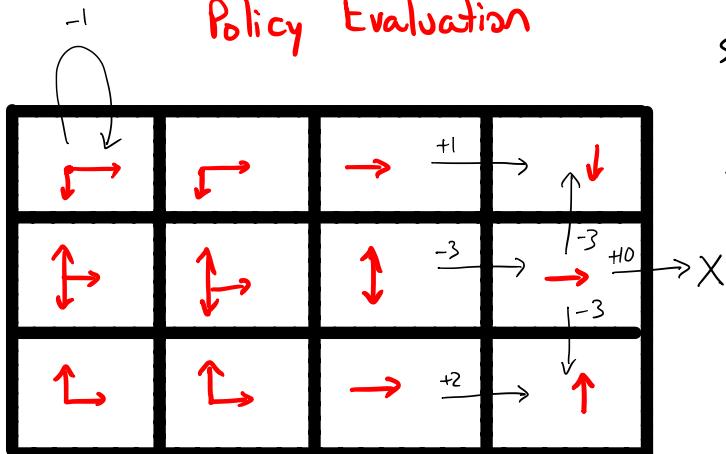
Stand  
Clap  
Wave

0	0	0	0
0	0	0	0
0	0	0	0

$$\forall s \quad v_0(s) = 0$$

$$v_{k+1}(s) = \sum_a \pi(a|s) \sum_{s', r} p(s', r | s, a) [r + \gamma v_k(s')]$$

## Policy Evaluation



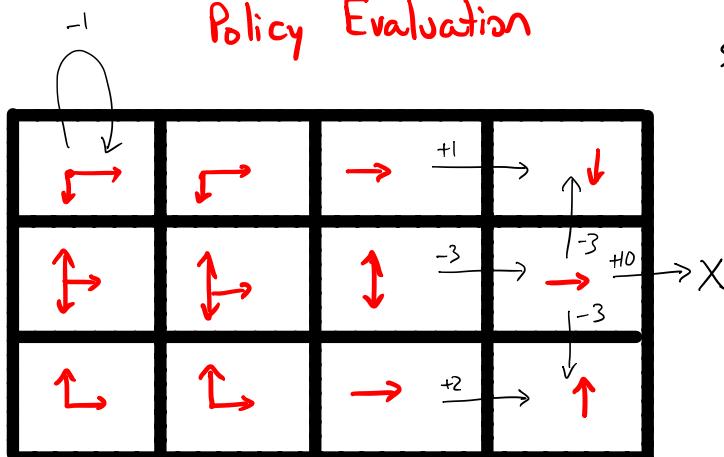
Stand  
Clap  
Wave

0,0	0,0	0,1	0,0
0,0	0,0	0,0	0,10
0,0	0,0	0,2	0,0

$$\forall s \quad v_0(s) = 0$$

$$v_{k+1}(s) = \sum_a \pi(a|s) \sum_{s'} p(s', r | s, a) [r + \gamma v_k(s')]$$

## Policy Evaluation



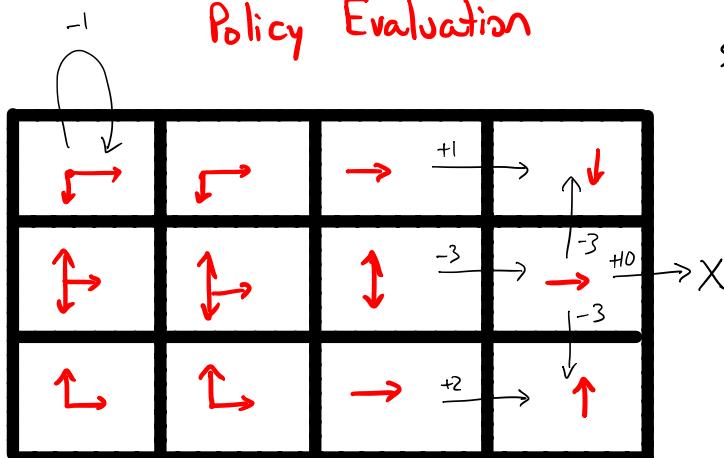
Stand  
Clap  
Wave

0,0,0	0,0,1,5	0,1,1,11	0,0,10
0,0,0	0,0,0	0,0,15 ?	0,10,10
0,0,0	0,0,1,1	0,2,2,12	0,0,10

$$\forall s \quad v_\pi(s) = 0$$

$$v_{k+1}(s) = \sum_a \pi(a|s) \sum_{s'} p(s', r | s, a) [r + \gamma v_k(s')]$$

## Policy Evaluation



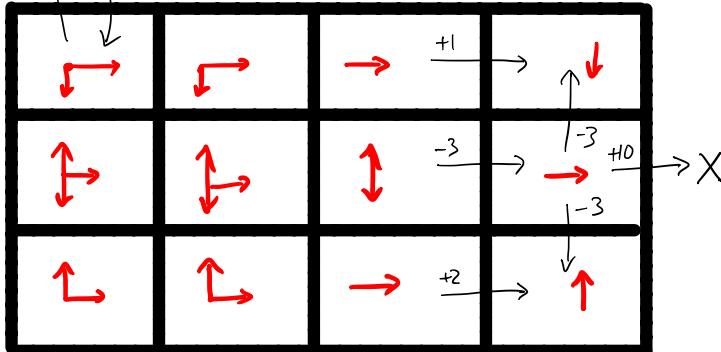
0,0,0	0,0,1,5 ?	0,1,1,11 11	0,0,10 10
0,0,0	0,0,0	0,0,1,5 11,5	0,1,0,10 10
0,0,0	0,0,1,1	0,2,2,12 12	0,0,10 10

Stand  
 Clap  
 Wave

$$\forall s \quad v_\pi(s) = 0$$

$$v_{k+1}(s) = \sum_a \pi(a|s) \sum_{s' r} p(s', r | s, a) [r + \gamma v_k(s')]$$

# Policy Evaluation

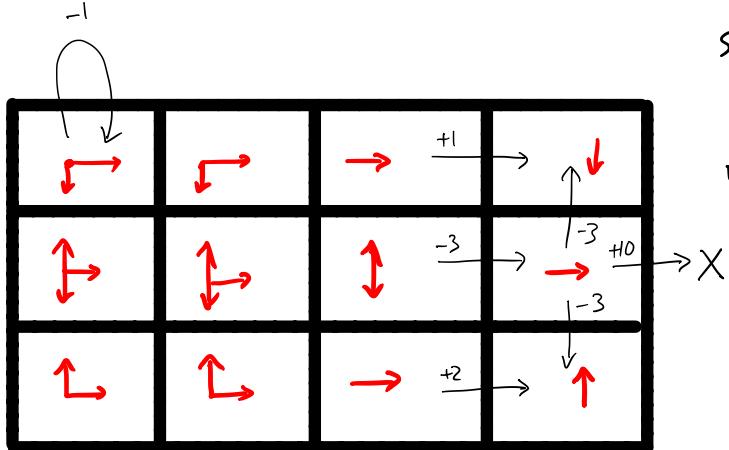


Stand  
Clap  
Wave

$0,0,0$ $\parallel \frac{3}{8}$	$0,0,1,5$ $\parallel \frac{1}{4}$	$0,1,1,11$ $\parallel 11$	$0,0,10$ $\parallel 10$
$0,0,0$ $\parallel \frac{1}{2}$	$0,0,0$ $\parallel \frac{1}{2}$	$0,0,15$ $\parallel 11.5$	$0,10,10$ $\parallel 10$
$0,0,0$ $\parallel \frac{3}{8}$	$0,0,1,1$ $\parallel \frac{3}{4}$	$0,2,2,12$ $\parallel 12$	$0,0,10$ $\parallel 10$

$$\text{As } v_0(s) = 0$$

$$v_{k+1}(s) = \sum_a \pi(a|s) \sum_{s' r} p(s', r | s, a) [r + \gamma v_k(s')]$$



Stand  
Clap  
Wave

$\rightarrow X$

Policy      Improvement

o	o	o	o
o	o	o	o
o	o	o	o

↑↑	↑↑	→	↓
↑↑	↑↑	→	↓
↑↑	↑↑	→	↓

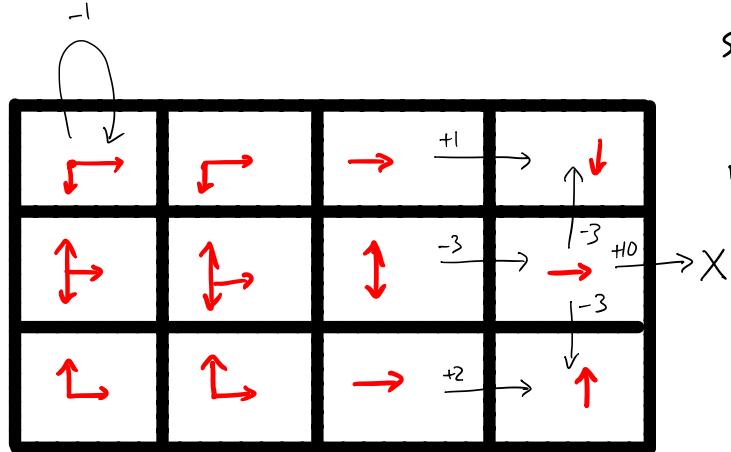
?
?
?









Stand  
Clap  
Wave

Policy Improvement

0	0	0	0
0	0	0	0
0	0	0	0

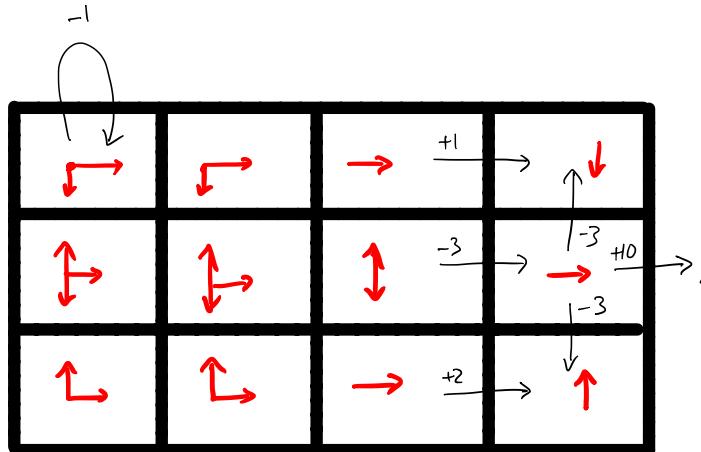

0	0	1	0
0	0	0	10
0	0	2	0









Stand  
Clap  
Wave

Policy

Improvement

0	0	0	0
0	0	0	0
0	0	0	0

↑	↑	→	↓
↑	↑	↓	→
↖	↖	→	↑

0	0	1	0
0	0	0	10
0	0	2	0

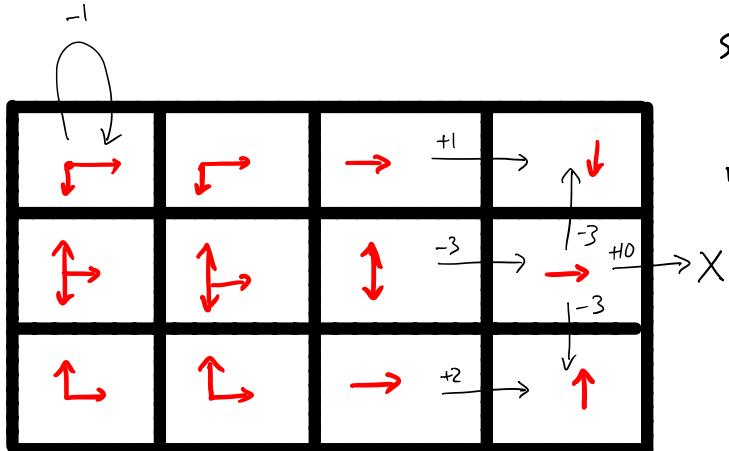
→		
	→	
		→

?







Stand  
Clap  
Wave

Policy

Improvement

0	0	0	0
0	0	0	0
0	0	0	0

↑	↑	→	↓
↑	↑	→	→
↑	↑	→	↑

0	0	1	0
0	0	0	10
0	0	2	0

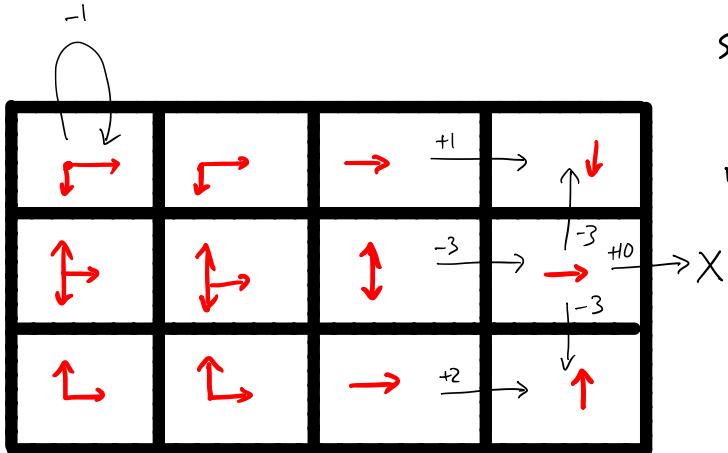
→			
		→	
		→	

0	1	1	10
0	0	7	10
0	2	2	10







Policy Improvement

0	0	0	0
0	0	0	0
0	0	0	0

0	0	1	0
0	0	0	10
0	0	2	0

0	1	1	10
0	0	7	10
0	2	2	10

?		
?		
?		

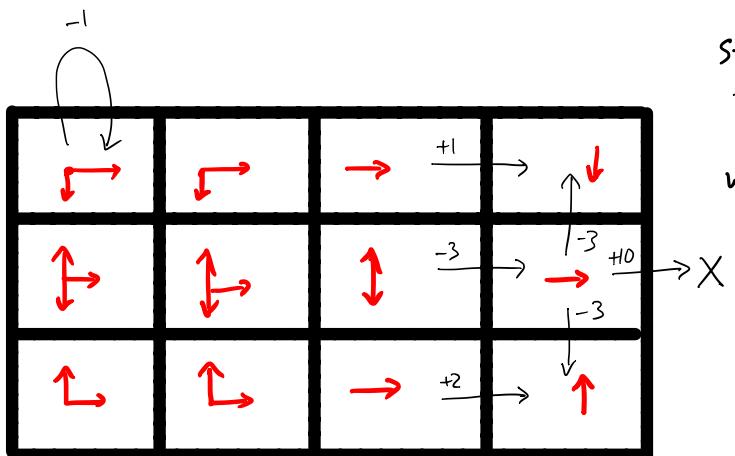


↑	↑	→	↓
↑	↑	↓	→
↑	↑	→	↑

→		
→		
→		

→	→	→	↓
↑	→	→	→
→	→	→	↑



Stand  
Clap  
Wave

Policy Improvement

0	0	0	0
0	0	0	0
0	0	0	0

↑	↑	→	↓
↑	↑	↓	→
↔	↔	→	↑

0	0	1	0
0	0	0	10
0	0	2	0

→			
		→	
		→	

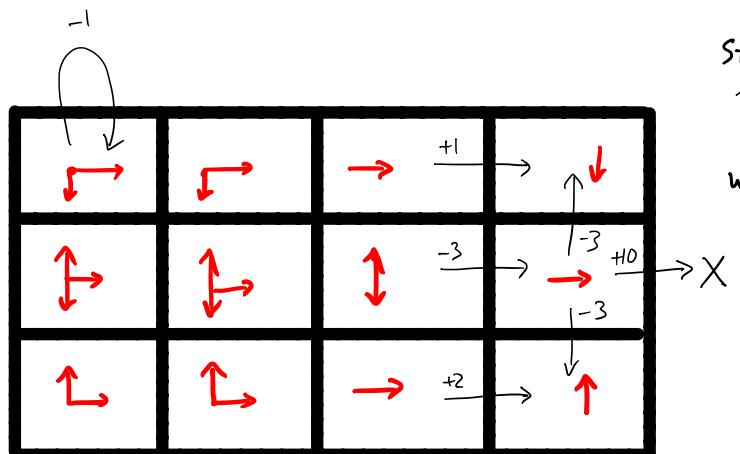
0	1	1	10
0	0	7	10
0	2	2	10

→	→	→	↓
↑	→	→	→
→	→	→	↑

1	1	11	10
0	7	7	10
2	2	12	10

?			





Policy Improvement

0	0	0	0
0	0	0	0
0	0	0	0

↗	↗	→	↓
↑	↗	↓	→
↖	↖	→	↑

0	0	1	0
0	0	0	10
0	0	2	0

→			
		→	
		→	

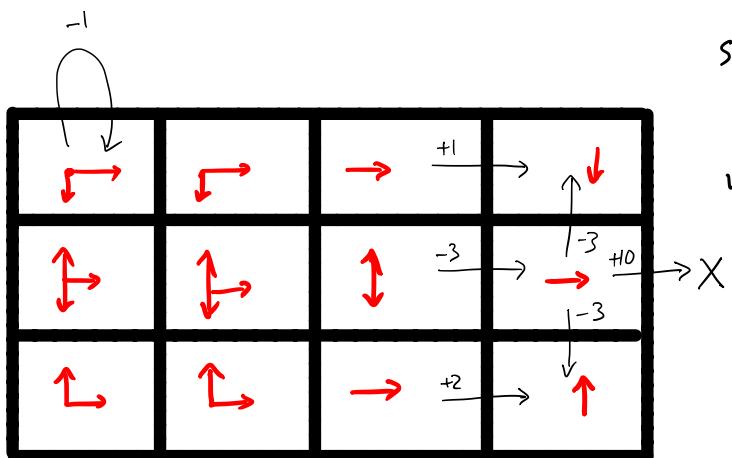
0	1	1	10
0	0	7	10
0	2	2	10

→	→	→	↓
↗	→	→	→
→	→	→	↑

1	1	11	10
0	7	7	10
2	2	12	10

→	→	→	↓
→	→	↓	→
→	→	→	↑

?			
?			

Stand  
Clap  
Wave

Policy

Improvement

0	0	0	0
0	0	0	0
0	0	0	0

$\uparrow \leftarrow$	$\uparrow \leftarrow$	$\rightarrow$	$\downarrow$
$\uparrow \leftarrow$	$\uparrow \leftarrow$	$\rightarrow$	$\uparrow$
$\uparrow \leftarrow$	$\uparrow \leftarrow$	$\rightarrow$	$\uparrow$

0	0	1	0
0	0	0	10
0	0	2	0

$\rightarrow$	$\rightarrow$	$\rightarrow$	$\downarrow$
$\uparrow \leftarrow$	$\uparrow \leftarrow$	$\rightarrow$	$\rightarrow$
$\rightarrow$	$\rightarrow$	$\rightarrow$	$\uparrow$

0	1	1	10
0	0	7	10
0	2	2	10

$\rightarrow$	$\rightarrow$	$\rightarrow$	$\downarrow$
$\uparrow \leftarrow$	$\uparrow \leftarrow$	$\rightarrow$	$\rightarrow$
$\rightarrow$	$\rightarrow$	$\rightarrow$	$\uparrow$

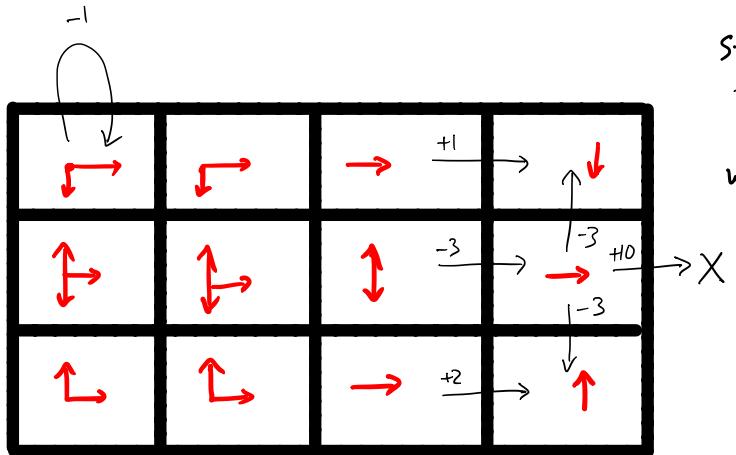
1	1	11	10
0	7	7	10
2	12	12	10

$\rightarrow$	$\rightarrow$	$\rightarrow$	$\downarrow$
$\rightarrow$	$\rightarrow$	$\downarrow$	$\rightarrow$
$\rightarrow$	$\rightarrow$	$\rightarrow$	$\uparrow$

1	11	11	10
7	7	12	10
2	12	12	10

$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$
$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$
$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$


?			



Stand  
Clap  
Wave

Policy

Improvement

0	0	0	0
0	0	0	0
0	0	0	0

↑	↓	→	↓
↑	↓	→	→
↑	↓	→	↑

0	0	1	0
0	0	0	10
0	0	2	0

→			
	→		
	→		↑

0	1	1	10
0	0	7	10
0	2	2	10

→	→	→	↓
↑	→	→	→
→	→	→	↑

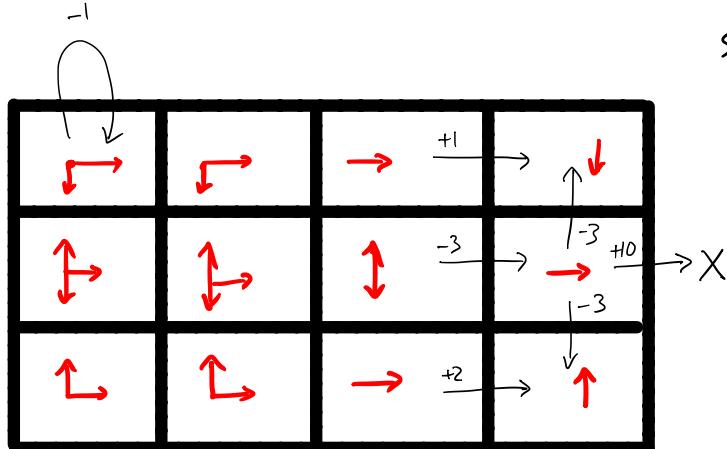
1	1	11	10
0	7	7	10
2	2	12	10

→	→	→	↓
→	→	↓	→
→	→	→	↑

1	11	11	10
7	7	12	10
2	12	12	10

→	↓	↓	↓
→	↓	↓	→
→	→	→	↑

$\pi^*$



Policy

Improvement

Stand  
Clap  
Wave

0	0	0	0
0	0	0	0
0	0	0	0

↗	↗	→	↓
↗	↗	→	→
↖	↖	→	↑

0	0	1	0
0	0	0	10
0	0	2	0

→			
	→		
		→	

0	1	1	10
0	0	7	10
0	2	2	10

→	→	→	↓
↗	↗	→	→
→	→	→	↑

1	1	11	10
0	7	7	10
2	2	12	10

→	→	→	↓
→	→	↓	→
→	→	→	↑

1	11	11	10
7	7	12	10
2	12	12	10

→	→	↓	↓
→	↗	↓	→
→	→	↑	↑

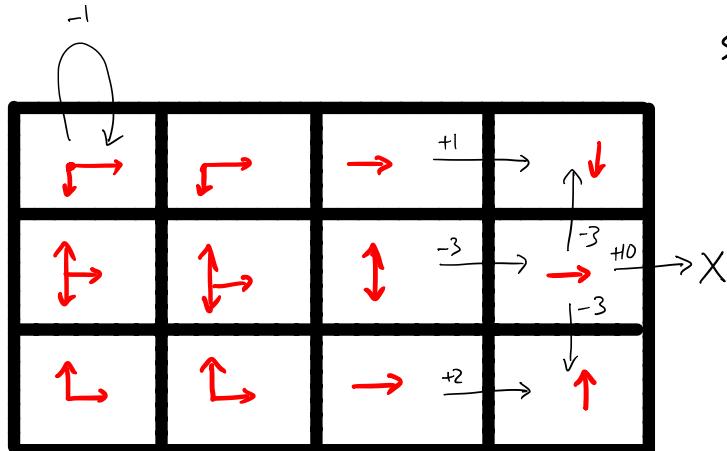
11	11	12	10
7	12	12	10
12	12	12	10

↓			
↗	↗	↓	
↗	↗	↑	

Value Iteration

$\pi^*$

$\pi^*$   
(also)



Stand  
Clap  
Wave

Policy

Improvement

0	0	0	0
0	0	0	0
0	0	0	0

↗	↗	→	↓
↗	↗	→	↓
↖	↖	→	↑

0	0	1	0
0	0	0	10
0	0	2	0

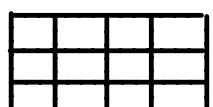
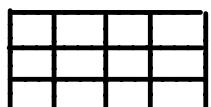
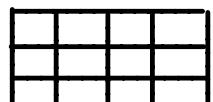
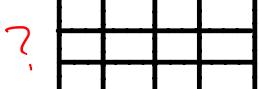
→			
		→	
		→	

0	1	1	10
0	0	7	10
0	2	2	10

→	→	→	↓
↗	↗	→	→
→	→	→	↑

1	1	11	10
0	7	7	10
2	2	12	10

→	→	↓	↓
→	↗	↓	→
→	→	↑	↑



Policy Iteration

1	11	11	10
7	7	12	10
2	12	12	10

→	→	↓	↓
→	↗	↓	→
→	→	↑	↑

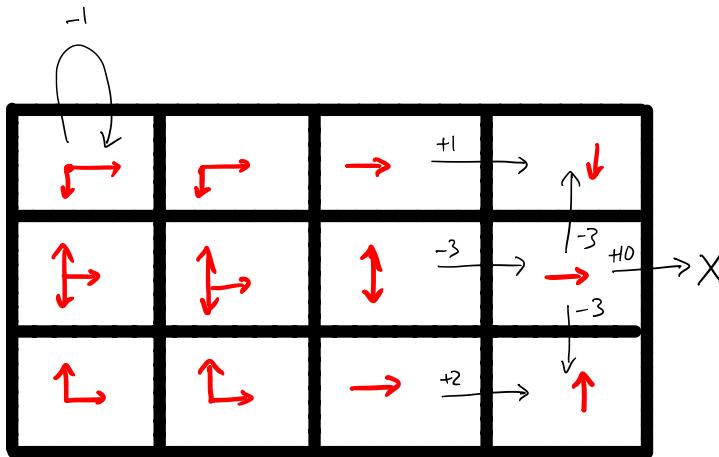
=  $\pi^*$

11	11	12	10
7	12	12	10
12	12	12	10

		↓	
↗	↗	↑	
↗	↗	↑	

=  $\pi^*$   
(also)

Value Iteration



Stand  
Clap  
Wave

Policy Improvement

0	0	0	0
0	0	0	0
0	0	0	0

↑	↑	→	↓
↑	↑	↓	→
↑	↑	↑	→

0	0	1	0
0	0	0	10
0	0	2	0

→	→	→	→
→	→	→	→
→	→	→	↑

0	1	1	10
0	0	7	10
0	2	2	10

→	→	→	↓
↑	→	→	→
→	→	→	↑

1	1	11	10
0	7	7	10
2	2	12	10

→	→	→	↓
→	↓	↓	→
→	→	↑	↑

11/8	11/4	11	10
11/2	11/2	11/2	10
11/8	11 3/4	12	10

?			

?			


Policy Iteration

1	11	11	10
7	7	12	10
2	12	12	10

→	→	↓	↓
→	↓	↓	→
→	→	↑	↑

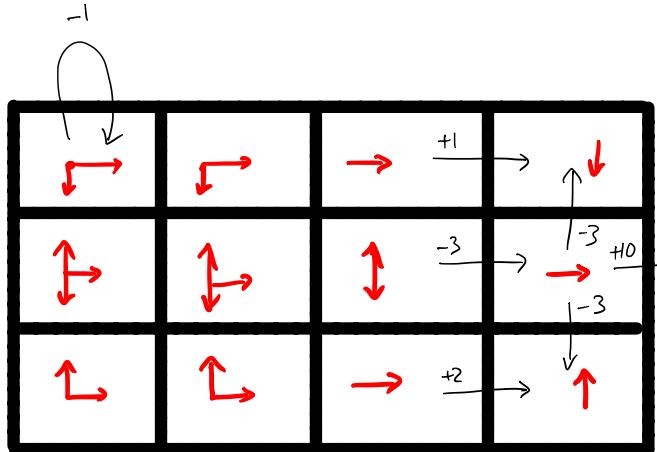
11	11	12	10
7	12	12	10
12	12	12	10

↓	↓	↑	↑
↑	↑	↑	↑
↑	↑	↑	↑

Value Iteration

=  $\pi^*$

=  $\pi^*$   
(also)



Policy      Improvement

Stand  
Clap  
Wave

0	0	0	0
0	0	0	0
0	0	0	0

$\uparrow$	$\downarrow$	$\rightarrow$	$\downarrow$
$\uparrow$	$\downarrow$	$\rightarrow$	$\rightarrow$
$\uparrow$	$\rightarrow$	$\rightarrow$	$\uparrow$

0	0	1	0
0	0	0	10
0	0	2	0

$\rightarrow$	$\rightarrow$	$\rightarrow$	$\downarrow$
$\uparrow$	$\rightarrow$	$\rightarrow$	$\rightarrow$
$\rightarrow$	$\rightarrow$	$\rightarrow$	$\uparrow$

0	1	1	10
0	0	7	10
0	2	2	10

$\rightarrow$	$\rightarrow$	$\rightarrow$	$\downarrow$
$\uparrow$	$\rightarrow$	$\downarrow$	$\rightarrow$
$\rightarrow$	$\rightarrow$	$\rightarrow$	$\uparrow$

1	1	11	10
0	7	7	10
2	2	12	10

$\rightarrow$	$\rightarrow$	$\downarrow$	$\downarrow$
$\rightarrow$	$\downarrow$	$\downarrow$	$\rightarrow$
$\rightarrow$	$\rightarrow$	$\uparrow$	$\uparrow$

1	11	11	10
7	7	12	10
2	12	12	10

$\rightarrow$	$\rightarrow$	$\downarrow$	$\downarrow$
$\rightarrow$	$\downarrow$	$\downarrow$	$\rightarrow$
$\rightarrow$	$\rightarrow$	$\uparrow$	$\uparrow$

$= \pi^*$

11	11	12	10
7	12	12	10
12	12	12	10

$\uparrow$	$\rightarrow$	$\downarrow$	
$\uparrow$	$\rightarrow$	$\downarrow$	
$\uparrow$	$\rightarrow$	$\uparrow$	

$= \pi^*$

(also)

Policy Iteration

Value Iteration

11 3/8	11 1/4	11	10
11 1/2	11 1/2	11 1/2	10
11 3/8	11 3/4	12	10

12	12	12	10
12	12	12	10
12	12	12	10

1	$\downarrow$	$\downarrow$	$\downarrow$
$\rightarrow$	$\downarrow$	$\downarrow$	$\rightarrow$
$\rightarrow$	$\rightarrow$	$\uparrow$	