

# Example: Temporal Difference Value Learning

States

	A	
B	C	D
	E	

Assume:  $\gamma = 1$ ,  $\alpha = 1/2$

$$-2 + 1 \cdot 0 = -2 \quad \boxed{r + \gamma V(s')}$$

$$V^\pi(B) \leftarrow (1 - \alpha) V^\pi(B) + \alpha (-2)$$

$$\leftarrow 0.5 \cdot 0 + \frac{1}{2} \cdot -2 = -1$$

Observed Transitions

$$-2 + 1 \cdot 8 = 6$$

$$V^\pi(C) \leftarrow 0.5 \cdot 0 + \frac{1}{2} \cdot 6 = 3$$

$$\boxed{\text{B, east, C, -2}}$$

$$\boxed{\text{C, east, D, -2}}$$

	0	
0	0	8
	0	

	0	
-1	0	8
	0	

	0	
-1	3	8
	0	

$$V^\pi(s) \leftarrow (1 - \alpha) V^\pi(s) + \alpha [R(s, \pi(s), s') + \gamma V^\pi(s')]$$