Reinforcement Learning

Kushal Shah @ Sitare University

Machine Learning

Machine Learning

Supervised Learning

Unsupervised Learning

Machine Learning

Supervised Learning

$$Y = f(X)$$

Unsupervised Learning

Machine Learning

Supervised Learning

$$Y = f(X)$$

Handwriting Recognition
Face Detection
Speech Recognition

Unsupervised Learning

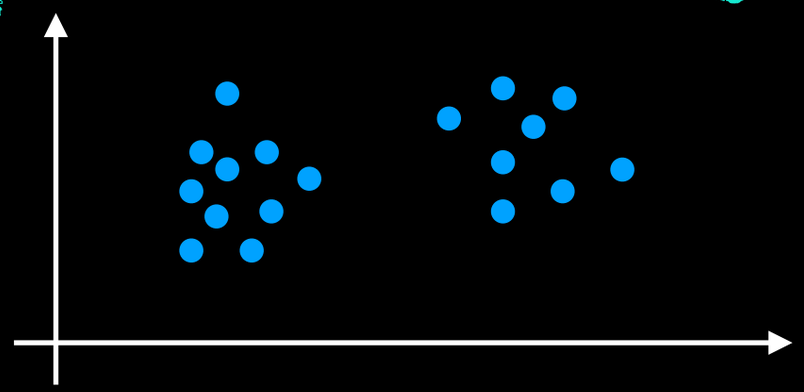
Machine Learning

Supervised Learning

$$Y = f(X)$$

Handwriting Recognition
Face Detection
Speech Recognition

Unsupervised Learning



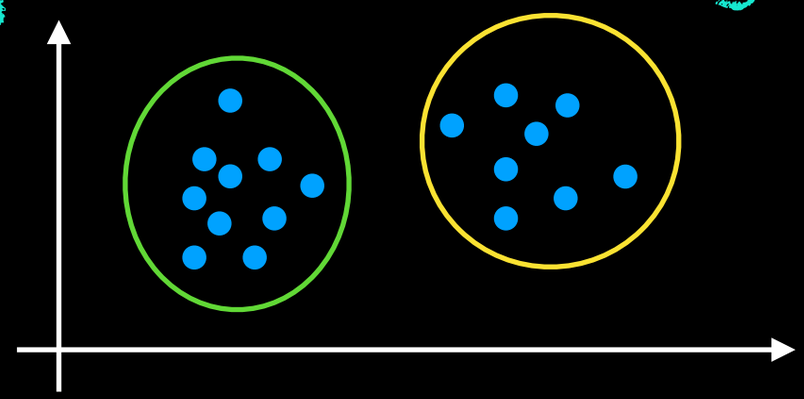
Machine Learning

Supervised Learning

$$Y = f(X)$$

Handwriting Recognition
Face Detection
Speech Recognition

Unsupervised Learning



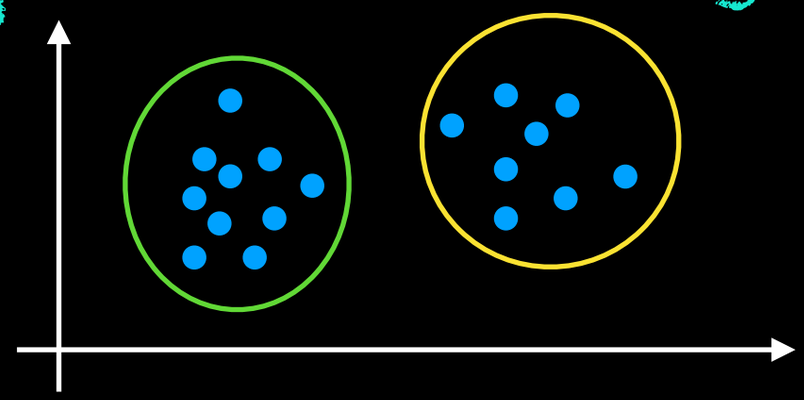
Machine Learning

Supervised Learning

$$Y = f(X)$$

Handwriting Recognition
Face Detection
Speech Recognition

Unsupervised Learning



Reinforcement Learning

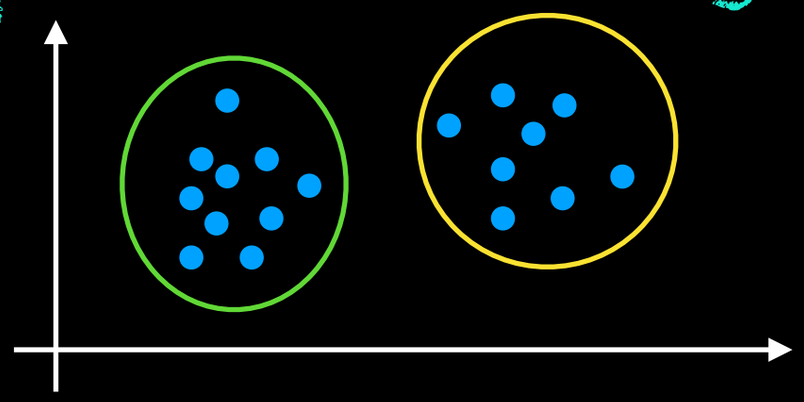
Machine Learning

Supervised Learning

$$Y = f(X)$$

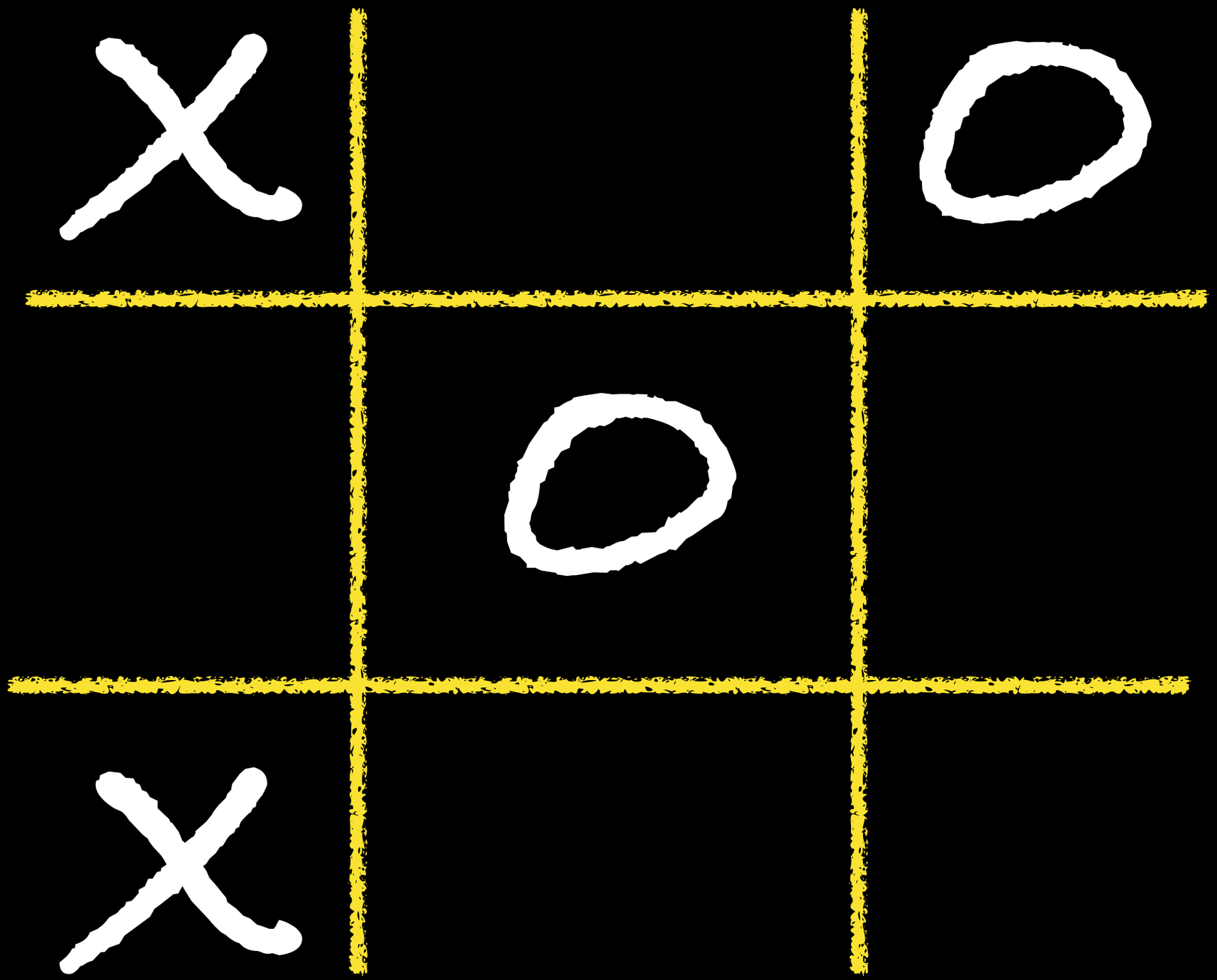
Handwriting Recognition
Face Detection
Speech Recognition

Unsupervised Learning

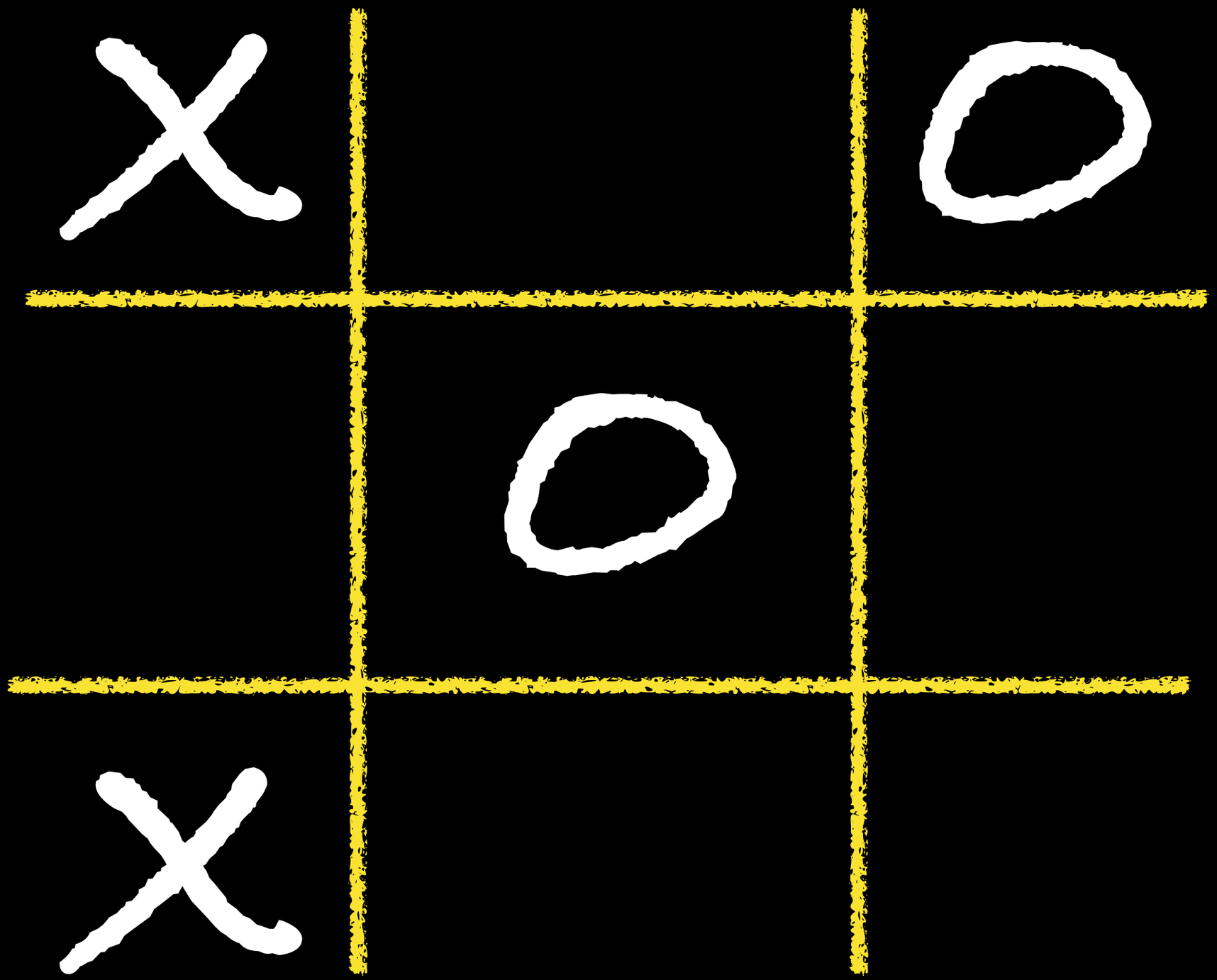


Reinforcement Learning

Learning to perform action in a certain environment,
closest to what biological systems do.

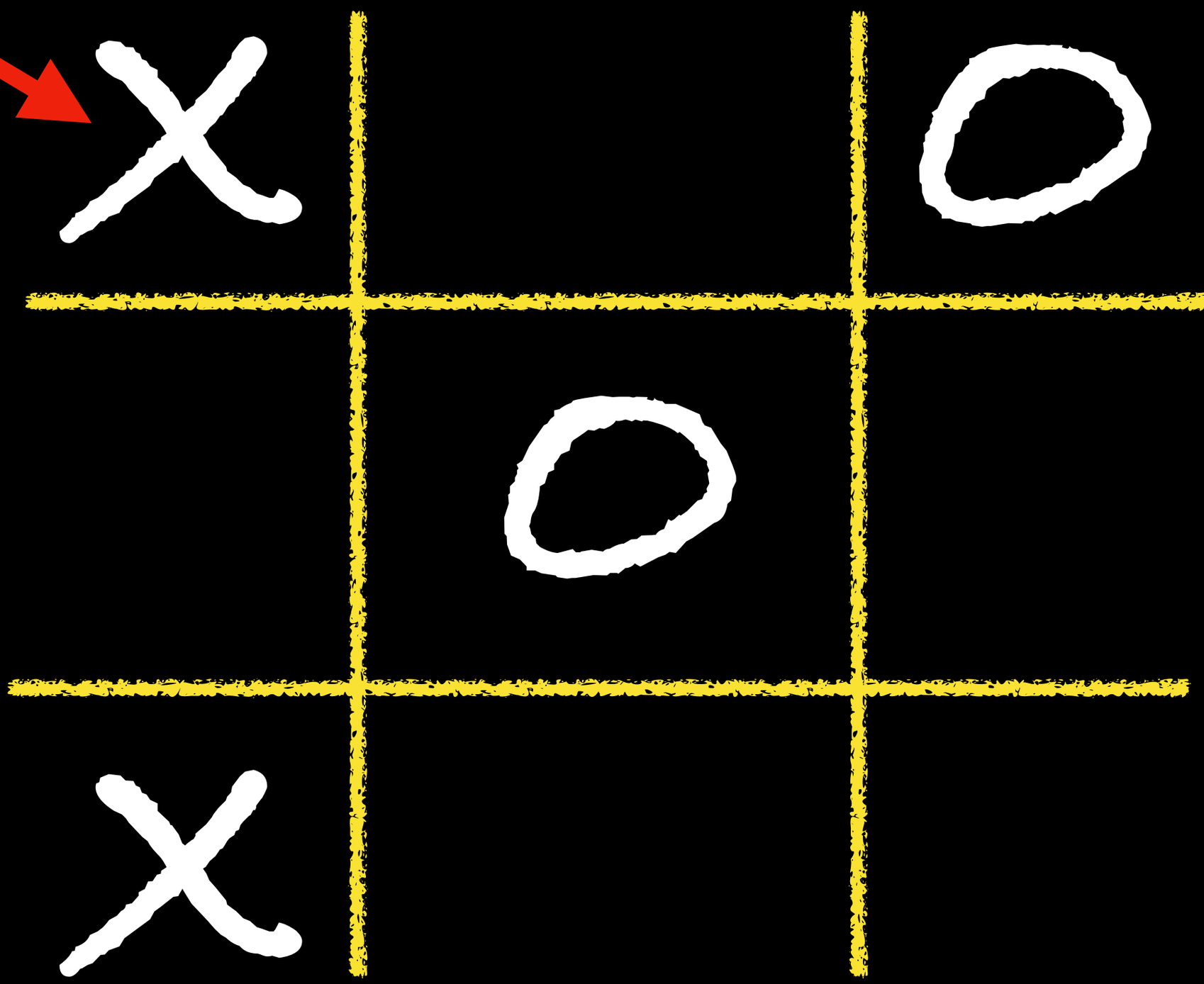


Environment



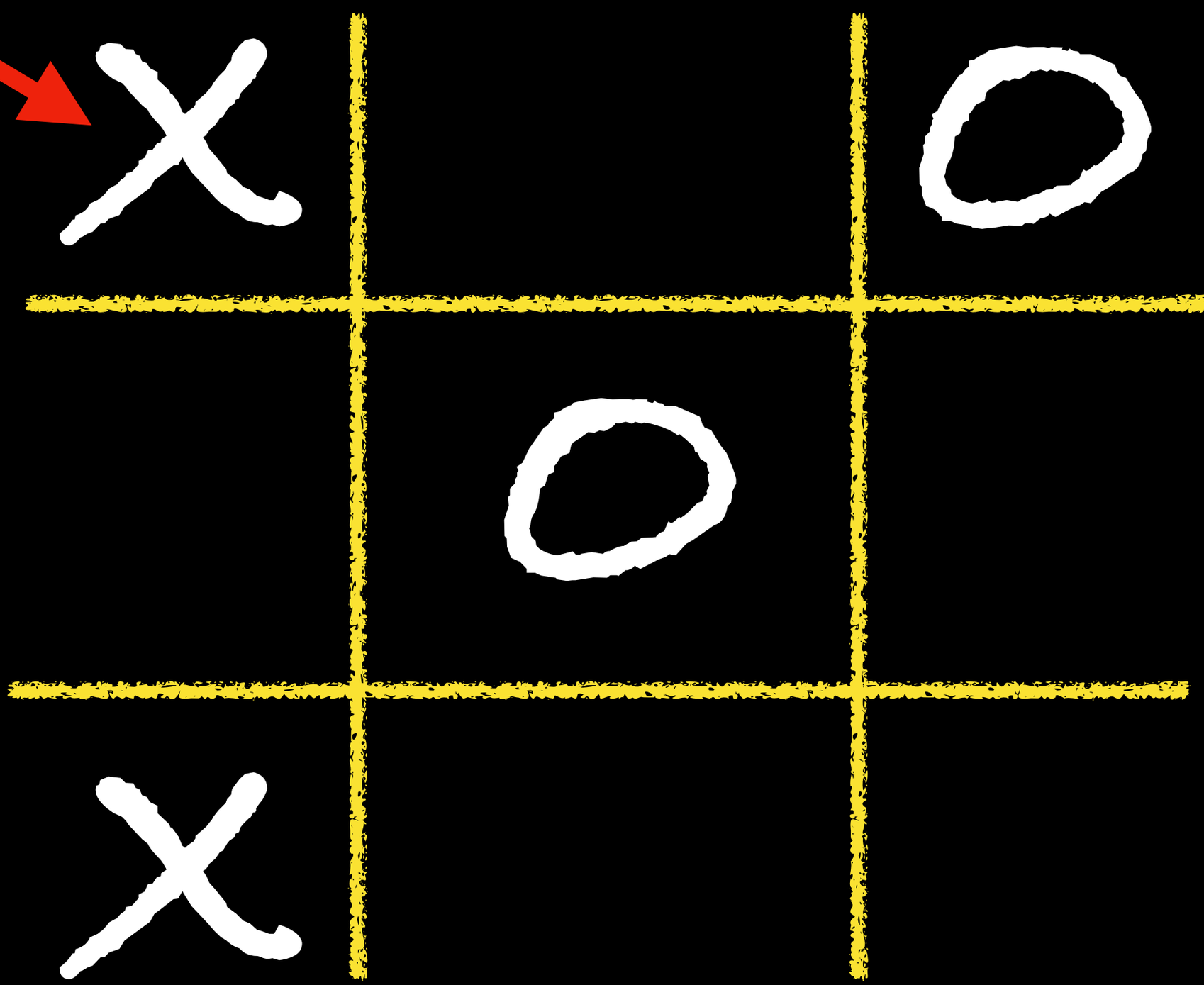
Agent

Environment



Agent

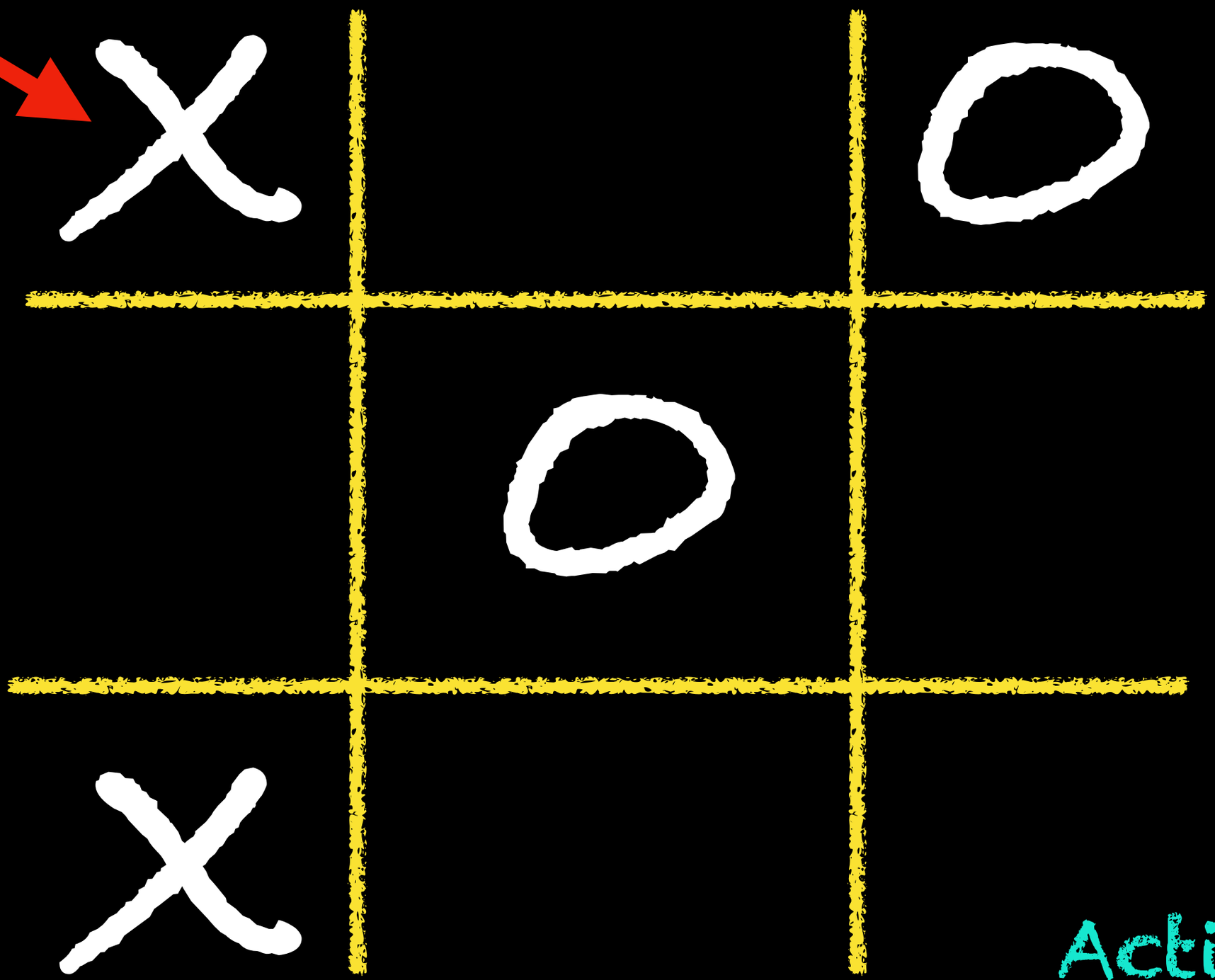
Environment



State/Configuration

Agent

Environment

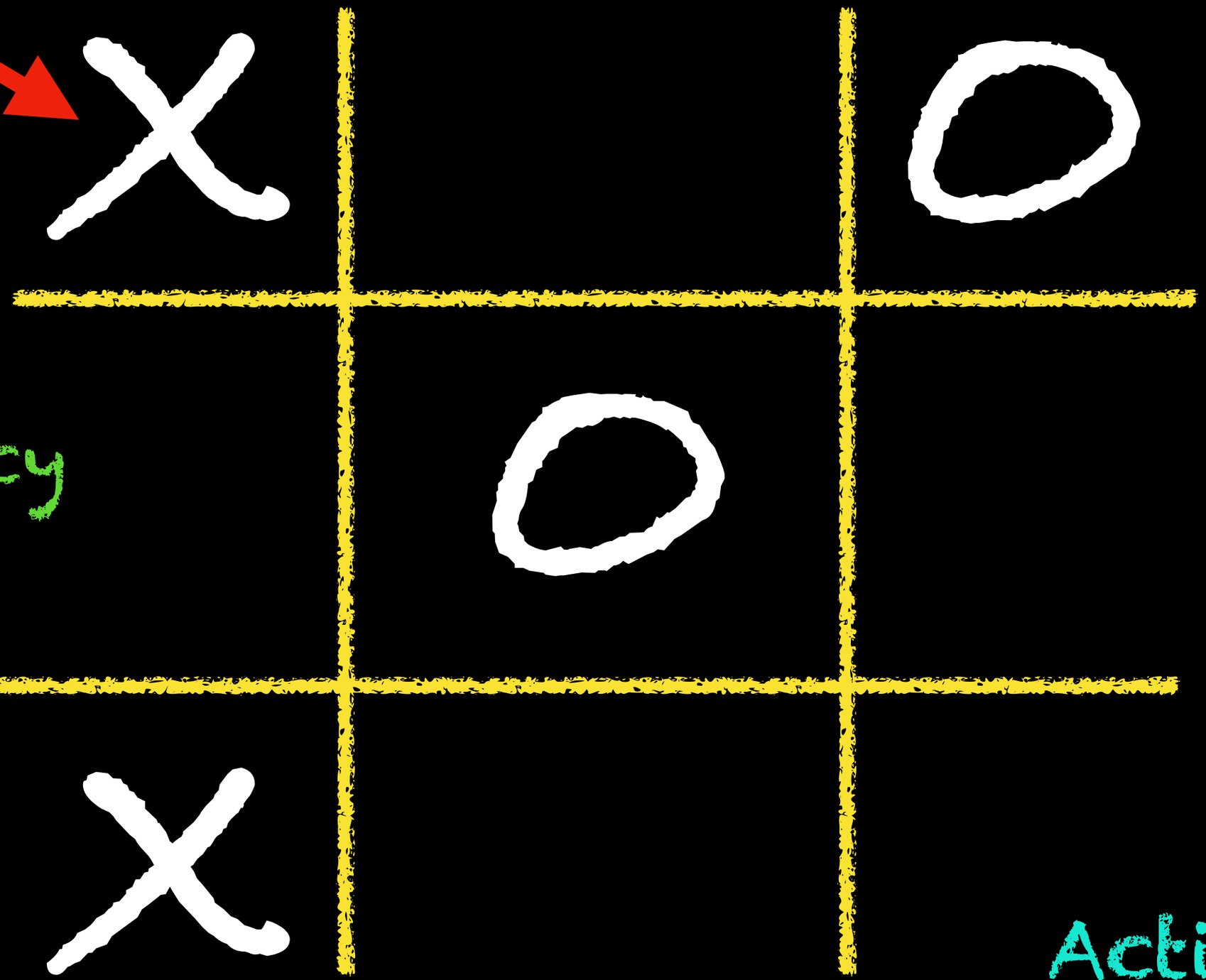


State/Configuration

Action

Agent

Environment



Policy

State/Configuration

Action

Agent

Environment

Policy

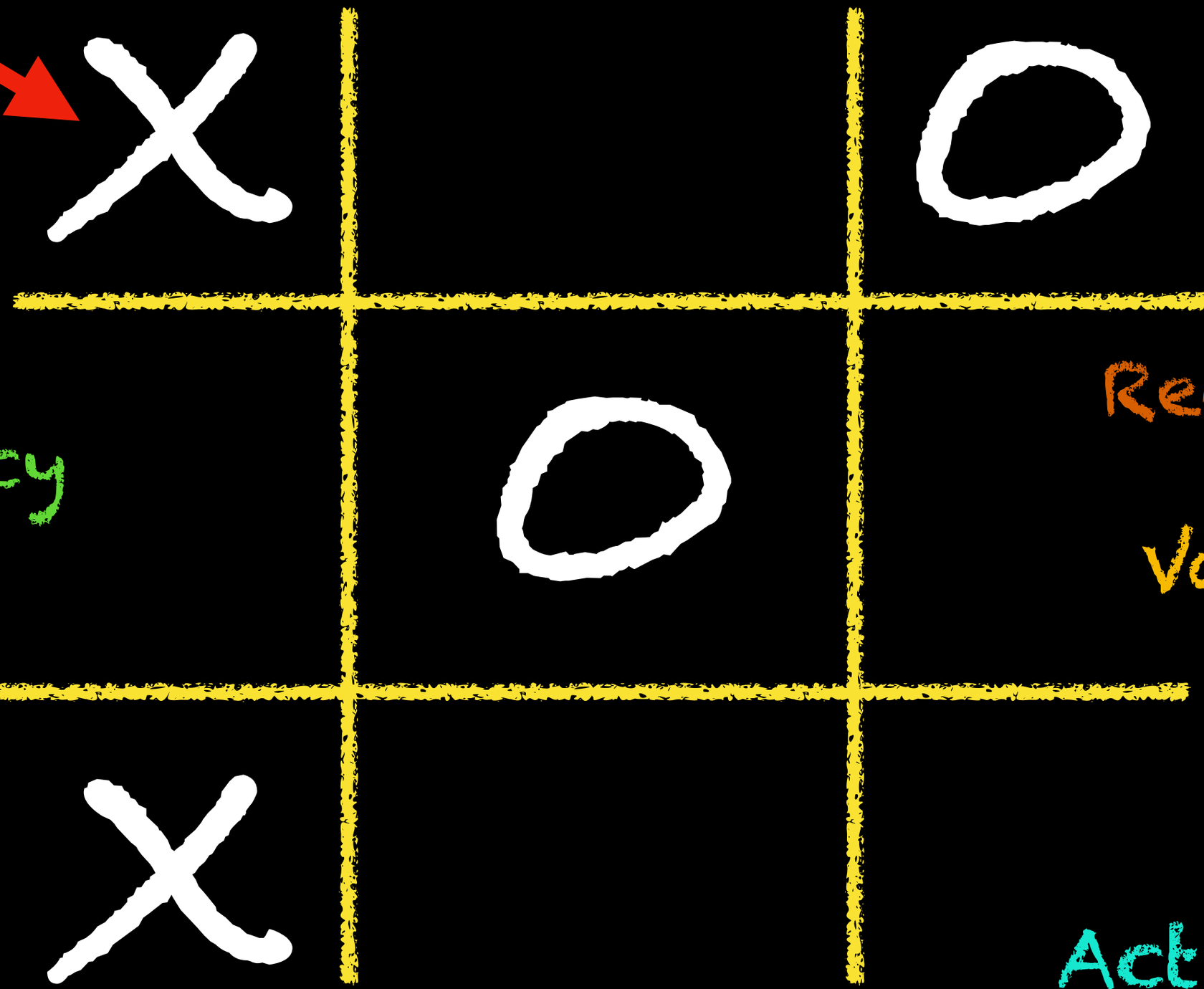
Reward

Action

State/Configuration

Agent

Environment



Policy

Reward

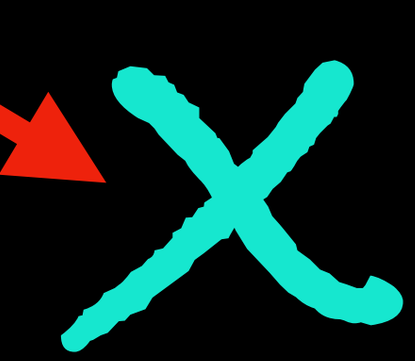
Value

Action

State/Configuration

Agent

Environment



Reward

Policy

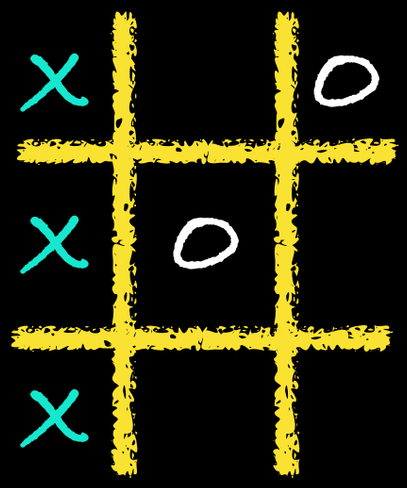
Value

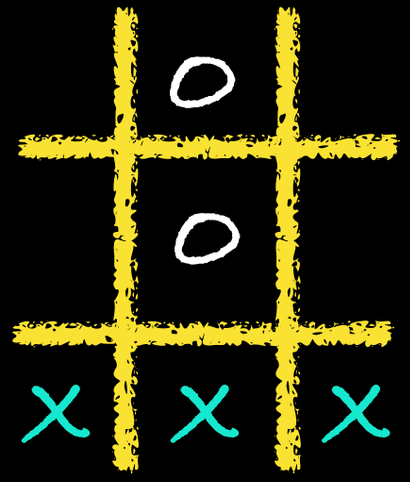
Action

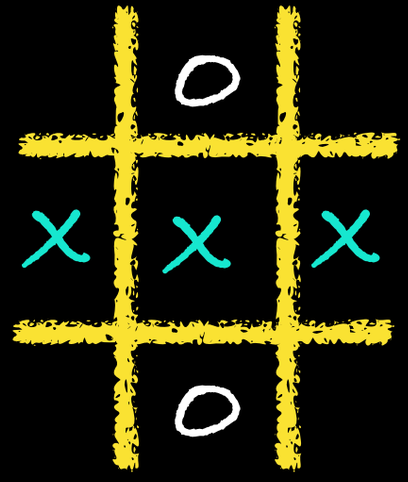
Total Number of States : 3^9

Win!

value = +1

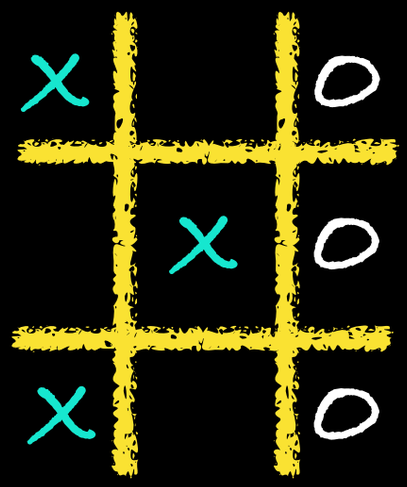


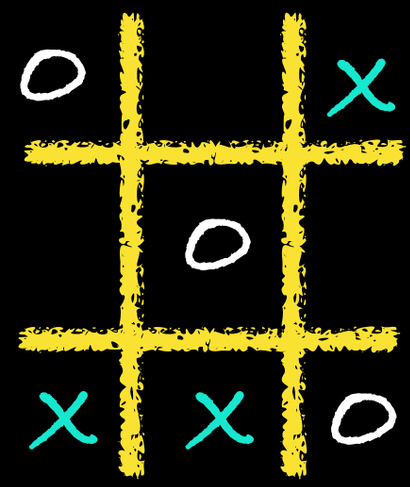


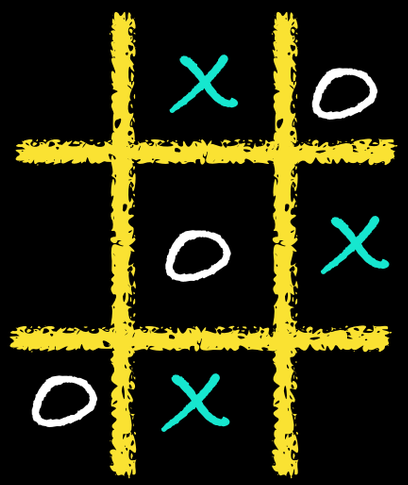


Lose!

value = -1

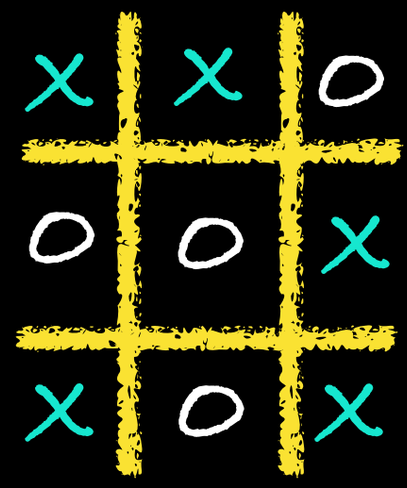


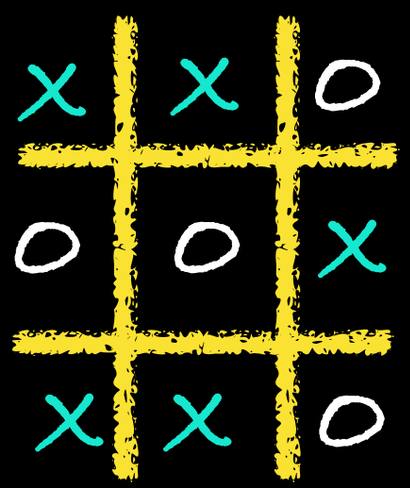


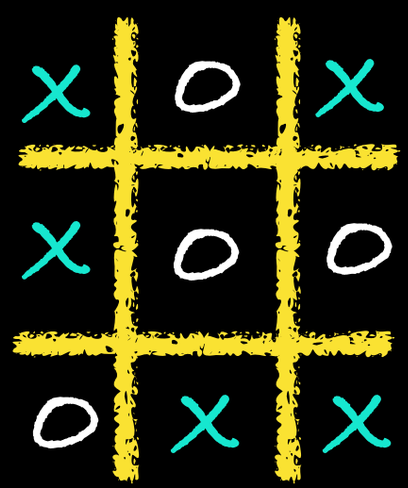


Draw!

value = 0







Total Number of States : 3^9

Agent

Policy

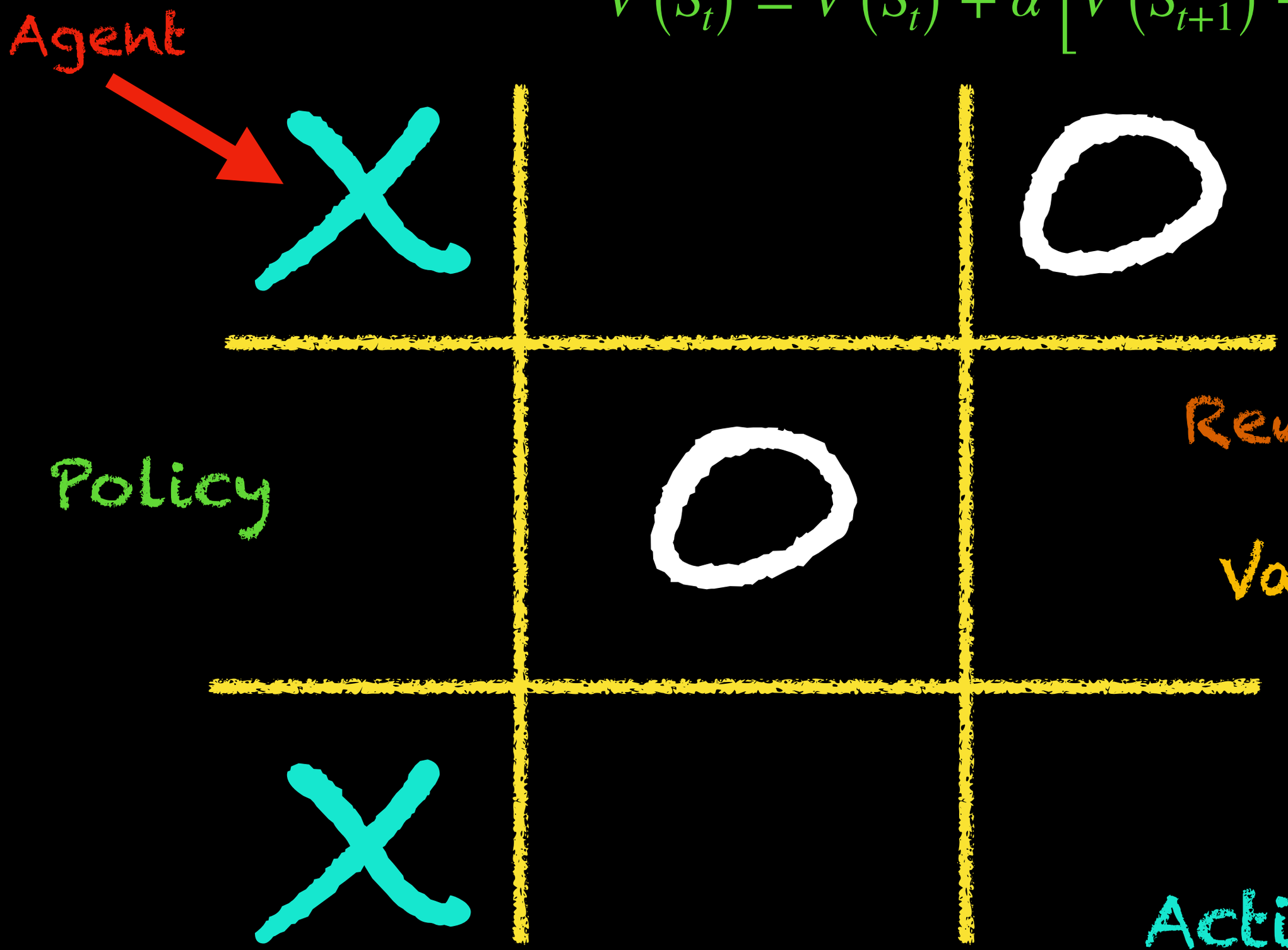
Reward

Value

Action

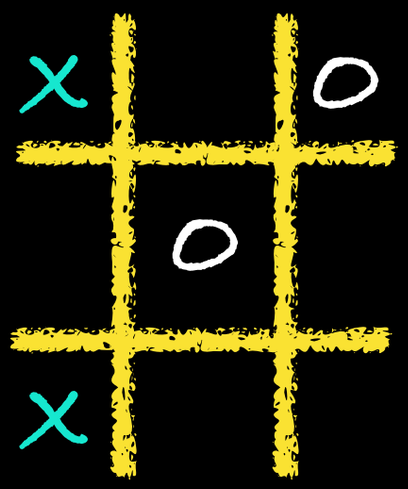
Total Number of States : 3^9

$$V(S_t) = V(S_t) + \alpha \left[V(S_{t+1}) - V(S_t) \right]$$



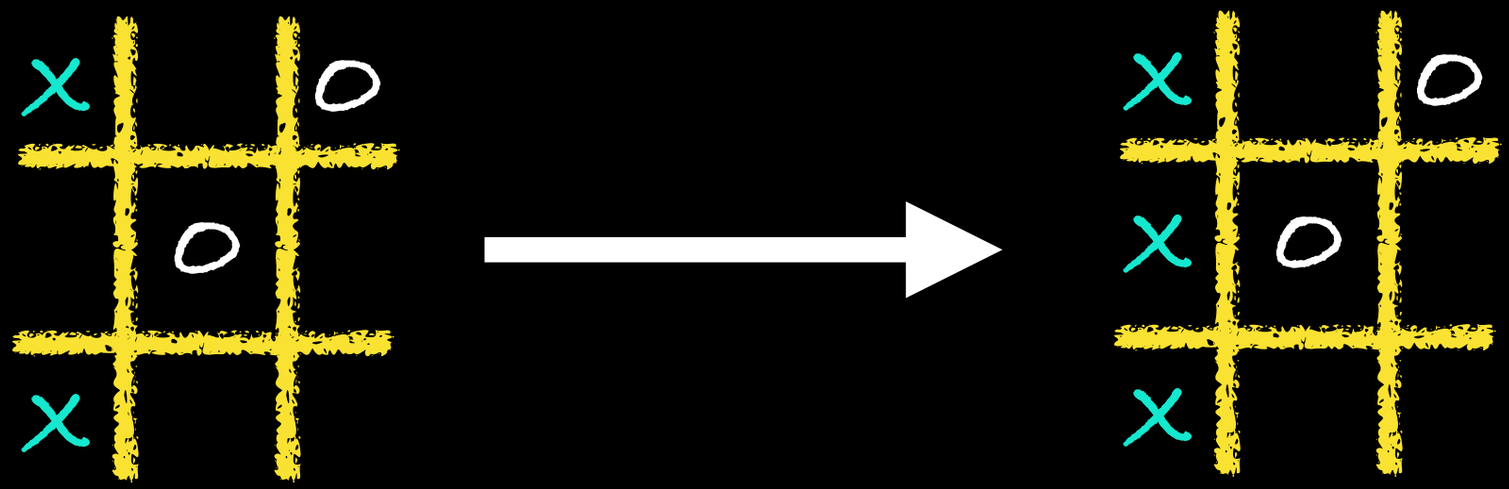
Total Number of States : 3^9

$$V(S_t) = V(S_t) + \alpha \left[V(S_{t+1}) - V(S_t) \right]$$



$$V(S_4) = 0$$

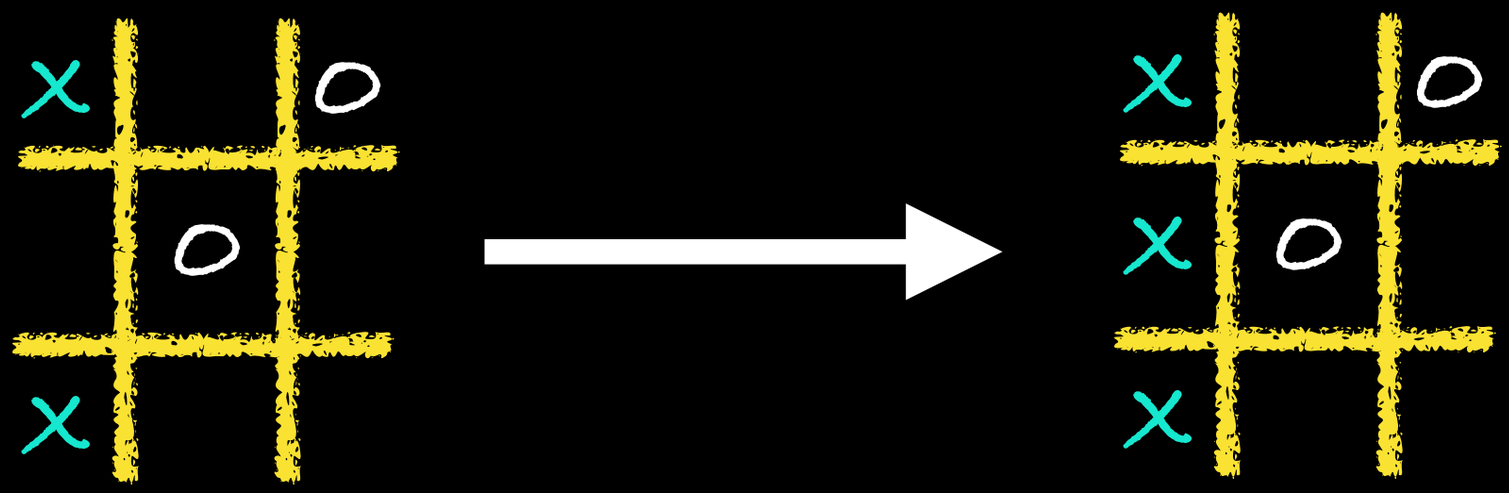
$$V(S_t) = V(S_t) + \alpha \left[V(S_{t+1}) - V(S_t) \right]$$



$V(S_4) = 0$

$V(S_5) = 1$

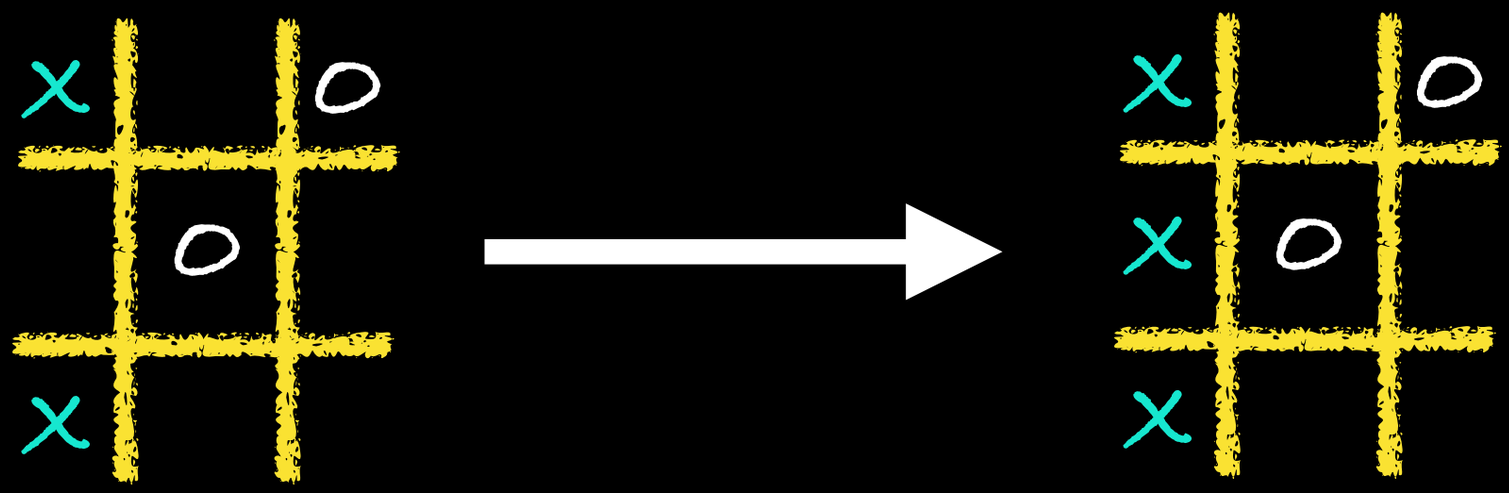
$$V(S_t) = V(S_t) + \alpha\left[V(S_{t+1}) - V(S_t)\right]$$



$V(S_4) = 0$ $V(S_5) = 1$

$$\rightarrow V(S_4) = 0 + \alpha\left[1 - 0\right] = \alpha$$

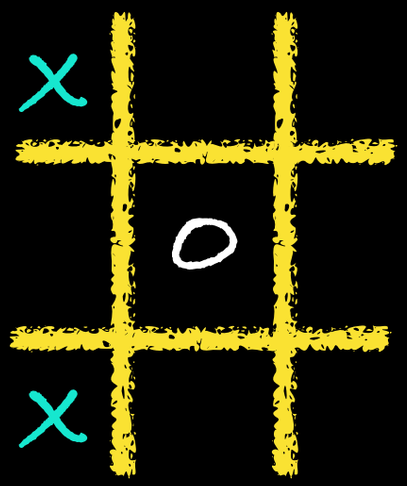
$$V(S_t) = V(S_t) + \alpha\left[V(S_{t+1}) - V(S_t)\right]$$



$V(S_4) = 0$

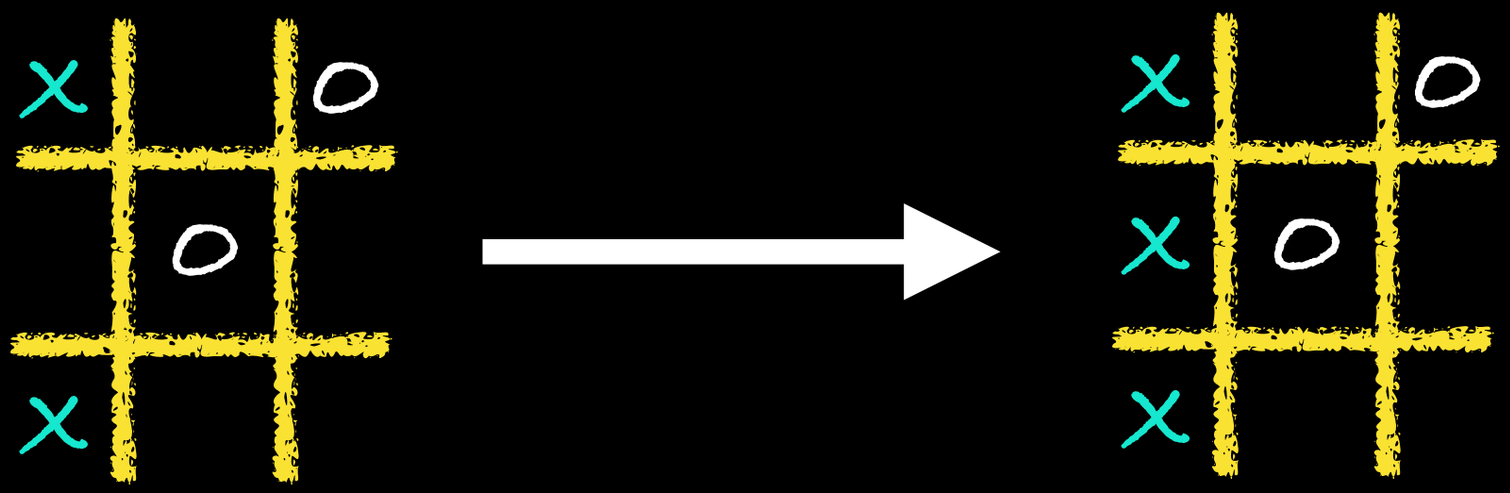
$V(S_5) = 1$

$$\rightarrow V(S_4) = 0 + \alpha\left[1 - 0\right] = \alpha$$



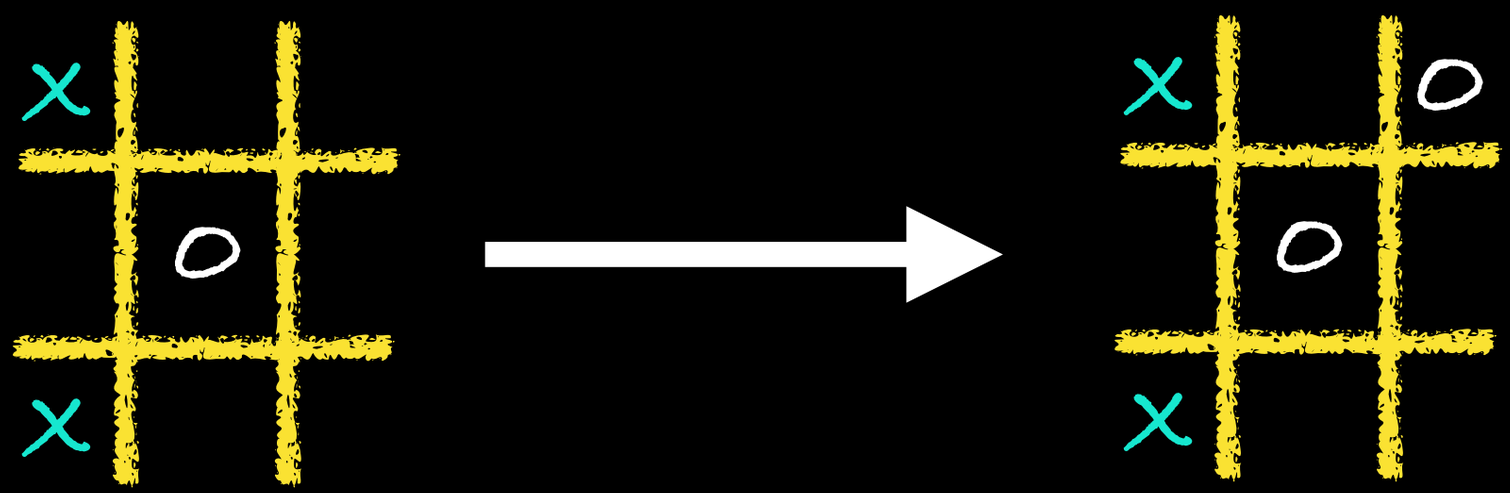
$V(S_3) = 0$

$$V(S_t) = V(S_t) + \alpha\left[V(S_{t+1}) - V(S_t)\right]$$



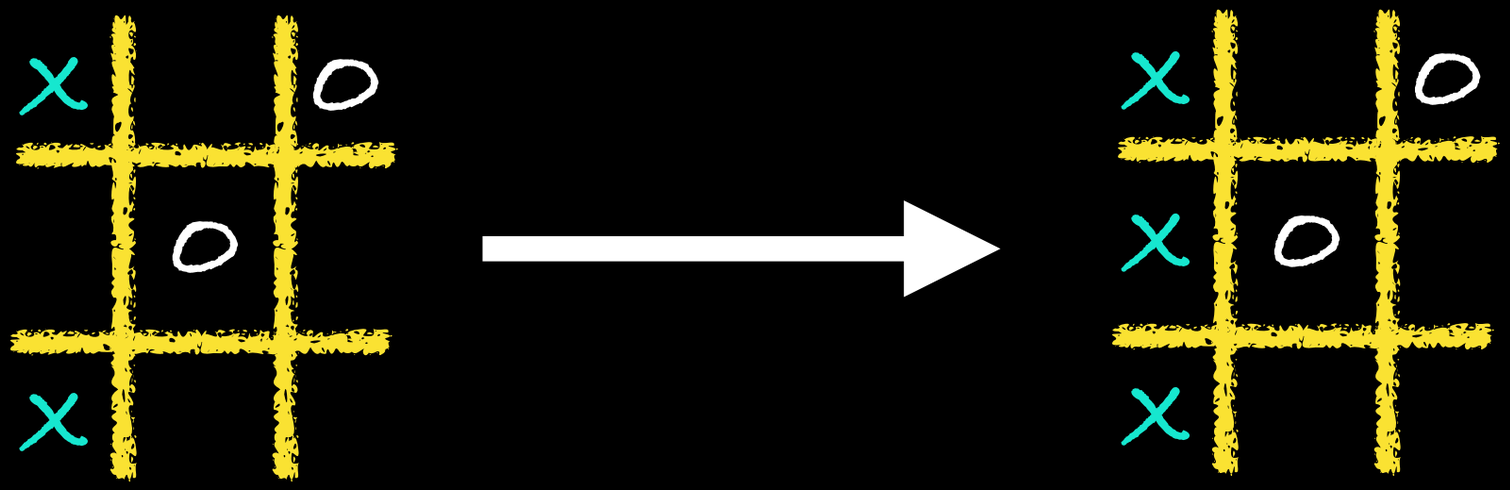
$V(S_4) = 0$ $\qquad$ $V(S_5) = 1$

$$\rightarrow V(S_4) = 0 + \alpha\left[1 - 0\right] = \alpha$$



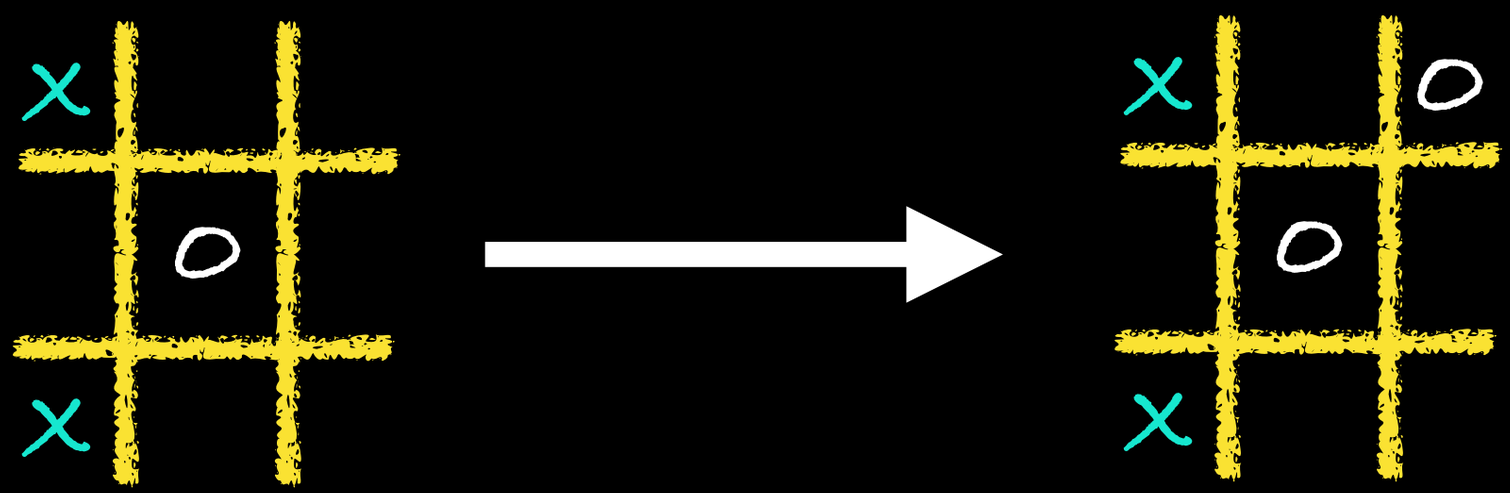
$V(S_3) = 0$ $\qquad$ $V(S_4) = \alpha$

$$V(S_t) = V(S_t) + \alpha\left[V(S_{t+1}) - V(S_t)\right]$$



$V(S_4) = 0$ $\qquad$ $V(S_5) = 1$

$$\rightarrow V(S_4) = 0 + \alpha\left[1 - 0\right] = \alpha$$



$V(S_3) = 0$ $\qquad$ $V(S_4) = \alpha$

$$\rightarrow V(S_3) = 0 + \alpha\left[\alpha - 0\right] = \alpha^2$$

$$V(S_t) = V(S_t) + \alpha\left[V(S_{t+1}) - V(S_t)\right]$$

Exploration

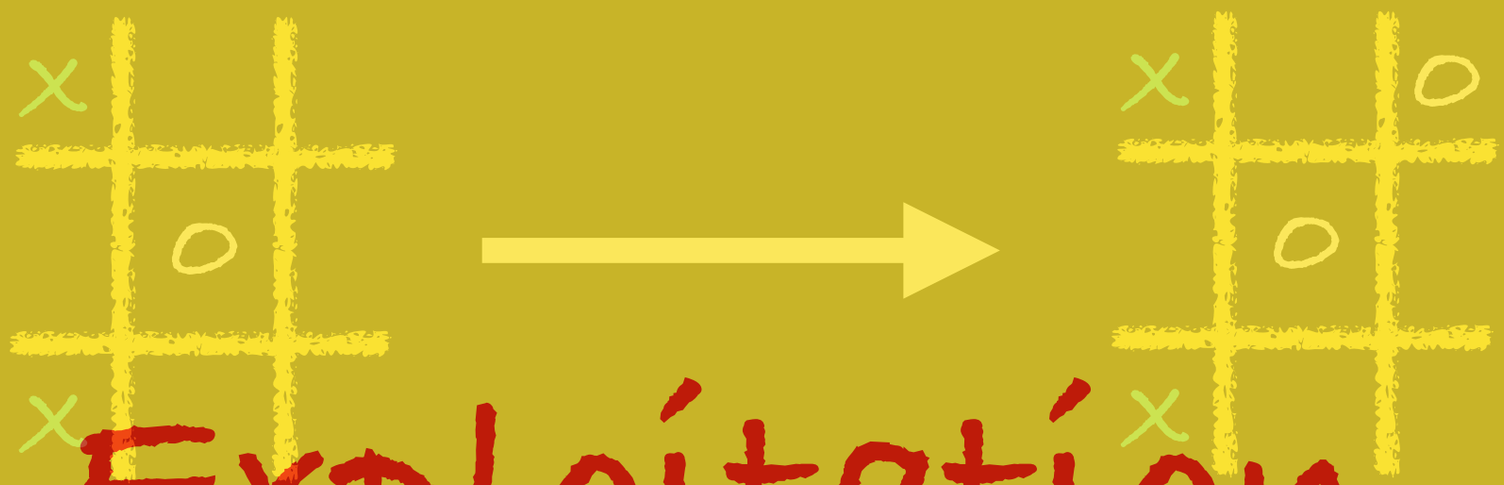
$V(S_4) = 0$

$V(S_5) = 1$

Policy

$\rightarrow V(S_4) = 0 + \alpha\left[1 - 0\right] = \alpha$

VS.

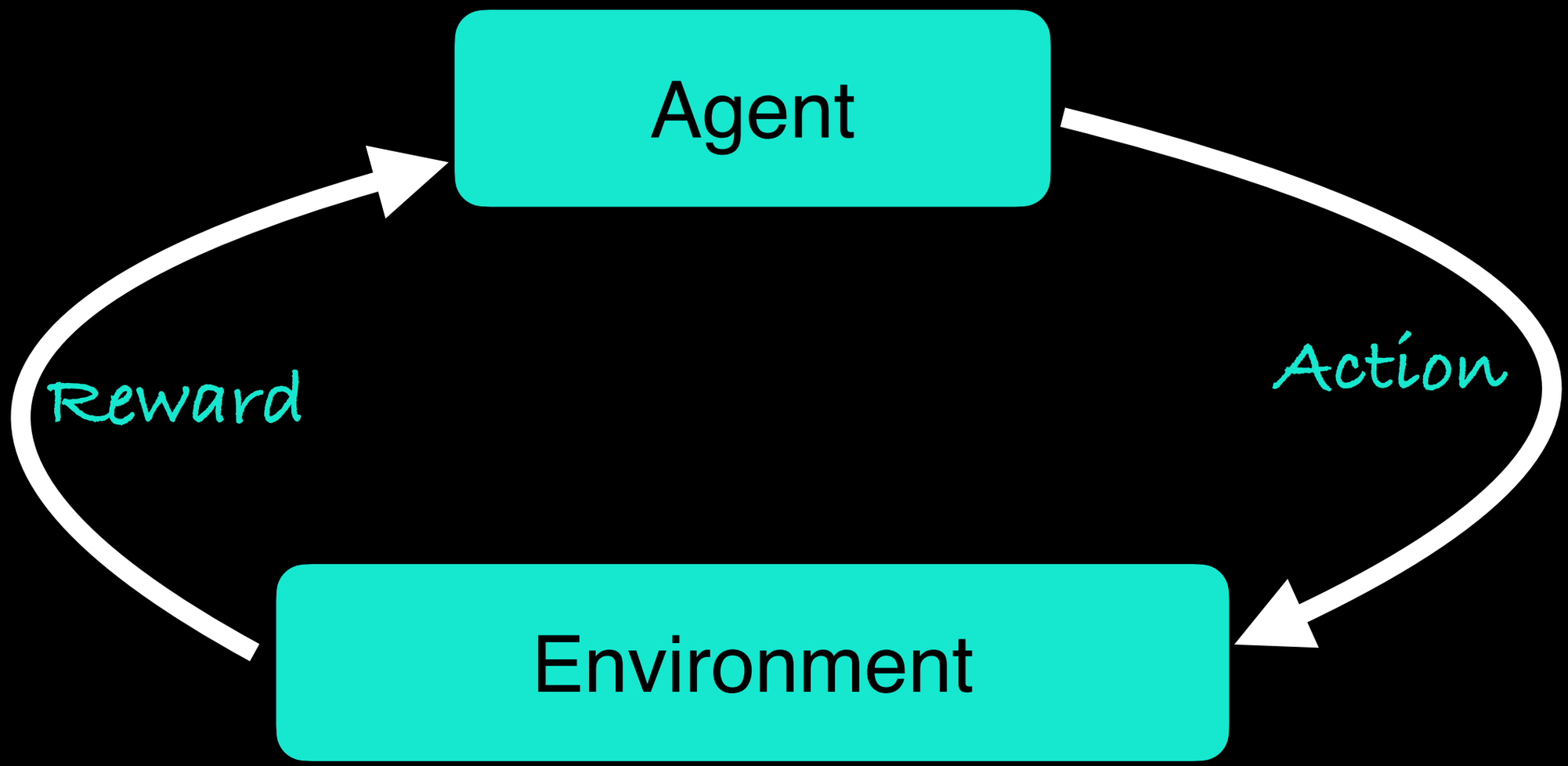


Exploitation

$V(S_3) = 0$

$V(S_4) = \alpha$

$\rightarrow V(S_3) = 0 + \alpha\left[\alpha - 0\right] = \alpha^2$

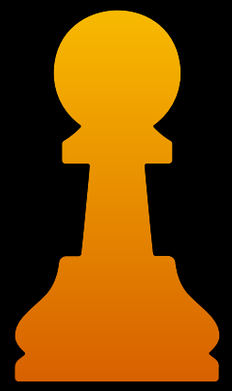

Agent
Reward
Action
Environment

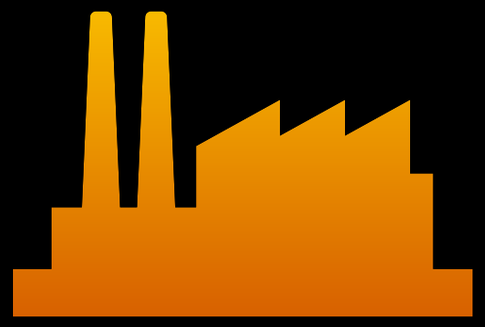
Agent

Reward

Action

Environment



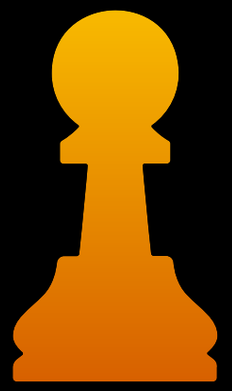


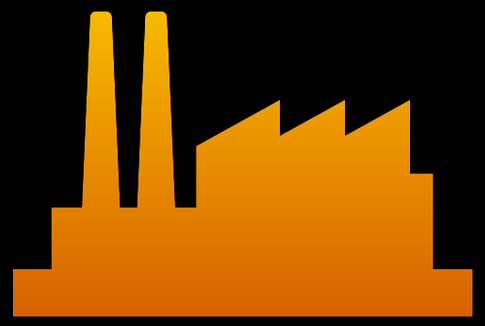
Agent

Reward

Action

Environment



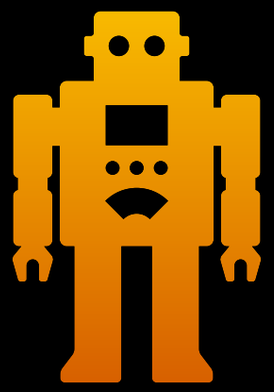


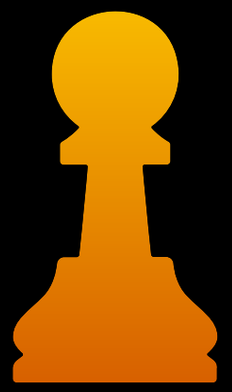
Agent

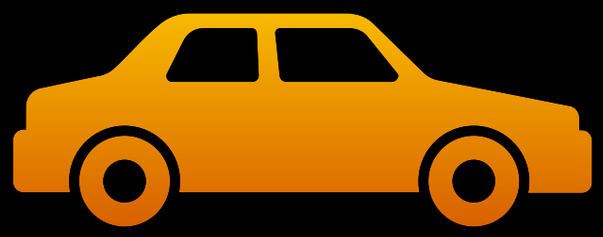
Reward

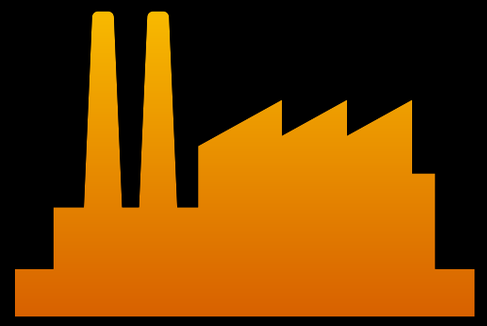
Action

Environment









Agent

Reward

Action

Environment