




Group.ai

202051131 - Pallav Sharma

202051132 - Parth Madan

202051133 - Apurv Patel

202051135 - Dev Patel



MENACE - Matchbox Educable Noughts and Crosses Engine



Learning Objective

1

State - Space Search

3

Markov Decision Process

2

Reinforcement Learning

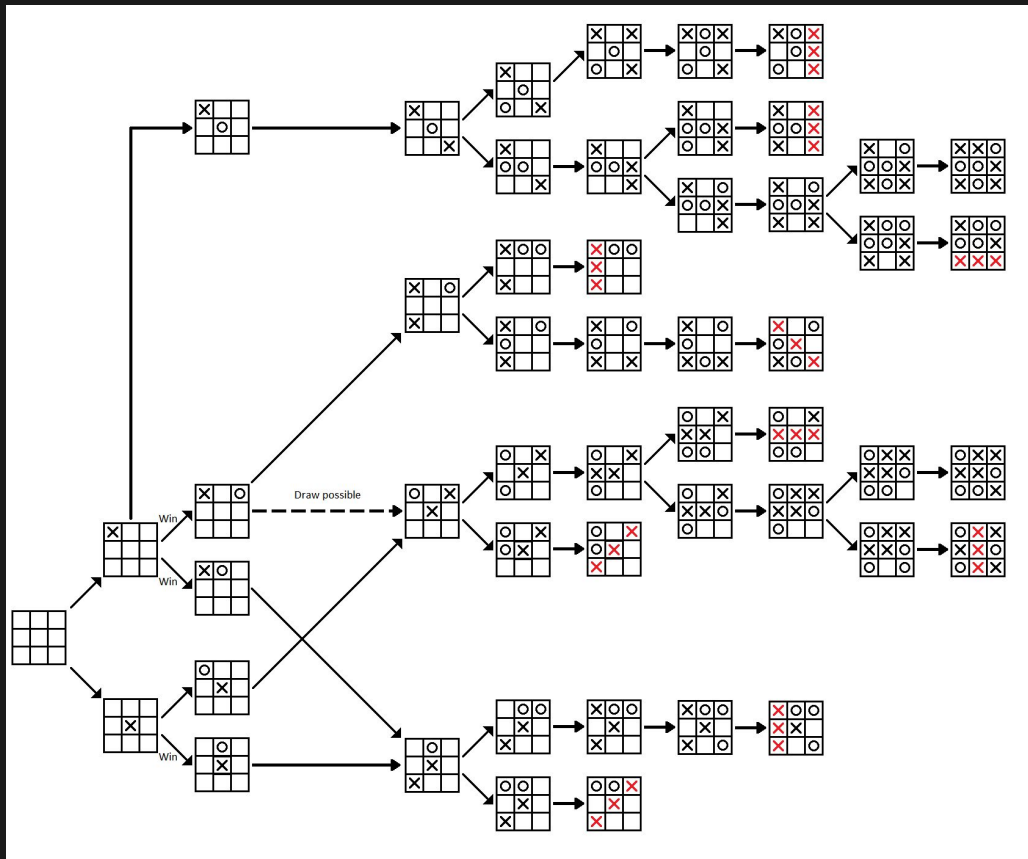
4

What is MENACE?



State - Space representation of Problem

1. All the states the system can be in are represented as nodes of the graph.
2. An action that can change the system from one state to another (e.g., a move in a game) is represented by a *link* from one node to another.
3. Links may be unidirectional (e.g., Tic-Tac-Toe) or bidirectional (e.g., Geographic move).
4. Search for a Solution.
5. It may be possible to reach the same state through many different paths.
6. There may be loops in the graph (can go round in a circle).



Representing Tic-Tac-Toe as State - Space Problem




Reinforcement Learning

Reinforcement learning is a machine learning training method based on rewarding desired behaviors and/or punishing undesired ones. In general, a reinforcement learning agent is able to perceive and interpret its environment, take actions and learn through trial and error.

Example

The problem is as follows: we have an agent and a reward, with many hurdles in between. The agent is supposed to find the best possible path to reach the reward. In our case we can say that the agent is us and the reward is to win the game.



The reinforcement learning process can be broken down into four main components:

Agent: The decision maker that interacts with the environment and takes actions.

Environment: The external system with which the agent interacts, and which provides the agent with information about its actions and the resulting states.

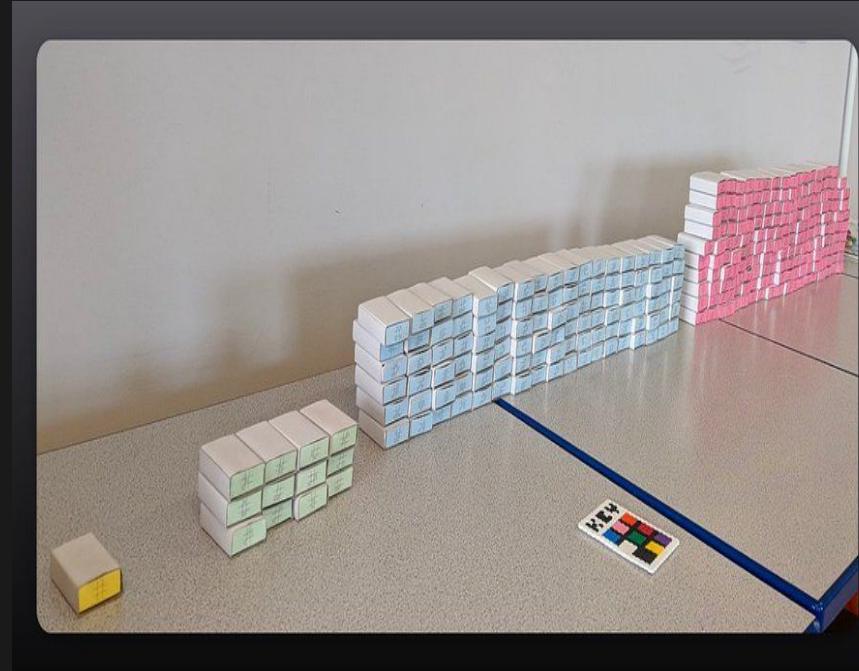
Actions: The actions that the agent can take in the environment. These actions may have different effects depending on the state of the environment.


Rewards: The feedback signal that the agent receives from the environment in response to its actions. The reward signal is used to guide the agent towards better decision making.

MENACE

In 1961, Donald Michie used 304 Matchboxes to design a solution that learned how to play tic-tac-toe.

He called his design MENACE (Matchbox Educable Noughts and Crosses Engine).





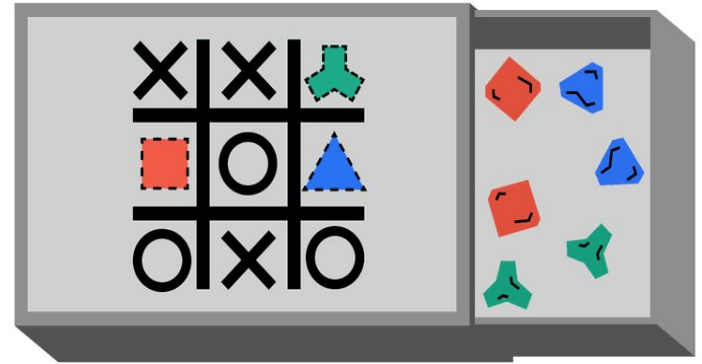
Michie didn't use explicit instructions to teach his system about tic-tac-toe.


Instead, the system learned to play through practice.

Each matchbox in 304 matchboxes printed different game state.

Inside Each matchboxes, there were coloured beads with different shapes. Every space on the board corresponded to one of these coloured shapes.

MENACE always plays first.





STAGE OF PLAY	NUMBER OF TIMES EACH COLOUR IS REPLICATED
1	4
3	3
5	2
7	1

Variation of the number of colour-replicates of a move according to the stage of play



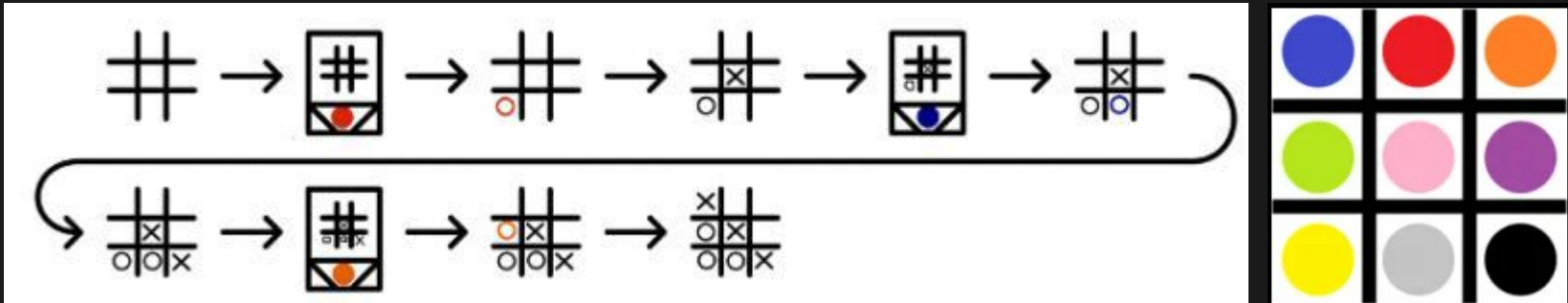
THE COLOR CODE USED IN THE MATCHBOX MACHINE


BY DONALD MICHIE

1 WHITE	2 LILAC	3 SILVER
8 BLACK	0 GOLD	4 GREEN
7 AMBER	6 RED	5 PINK

The front of each matchbox has a game of noughts and crosses printed on it. Each coloured bead inside the matchbox represents the next move MENACE could make.

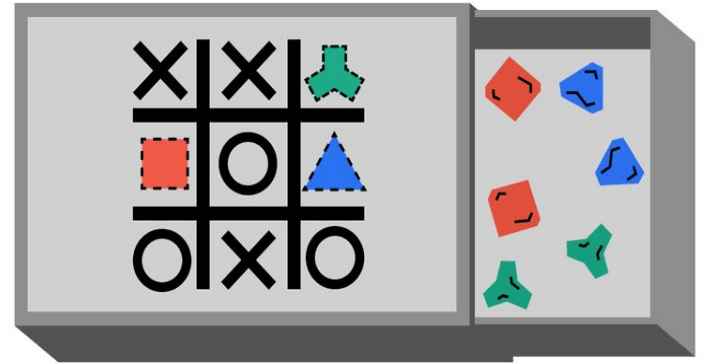
To find out what move MENACE wants to make locate the matchbox which matches the current state of your game. Then shake the box and take out a bead. MENACE plays in the position corresponding to the coloured bead came out.






For example, if the assistant drew a green bead, MENACE's marker was placed on the upper-right corner of the board.

Michie's revolutionary idea was to adjust the contents of the matchboxes at the end of each game.





He had only three rules:

First, if MENACE loses, throw away every bead played during that game.

This rule made it less likely for MENACE to play a losing move in the future.

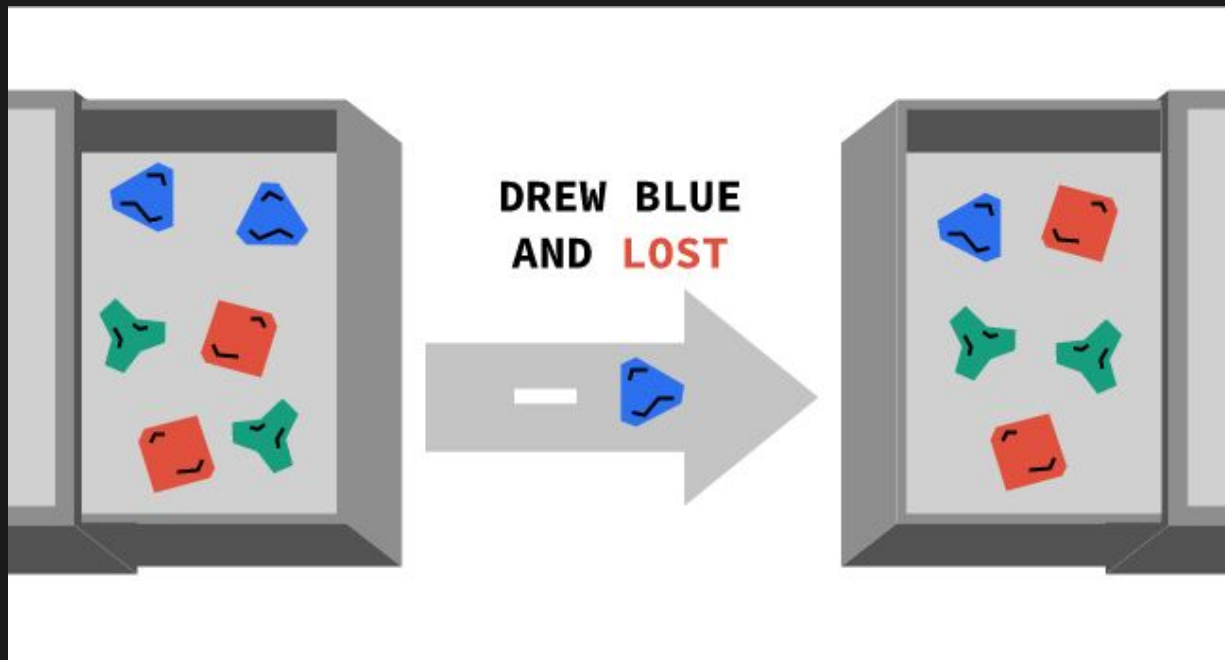
Second, if MENACE wins, put the beads drawn back in their matchboxes and add three more beads of the same color.

This rule made it more likely for MENACE to repeat the moves that led to the win.

Finally, the third rule was to return every bead to its box after a draw and add one more bead of the same color

Before the first game, MENACE was equally likely to play any move because every matchbox contained the same number of bead colors.

But with every game, MENACE made some plays more often than others.
MENACE started learning.

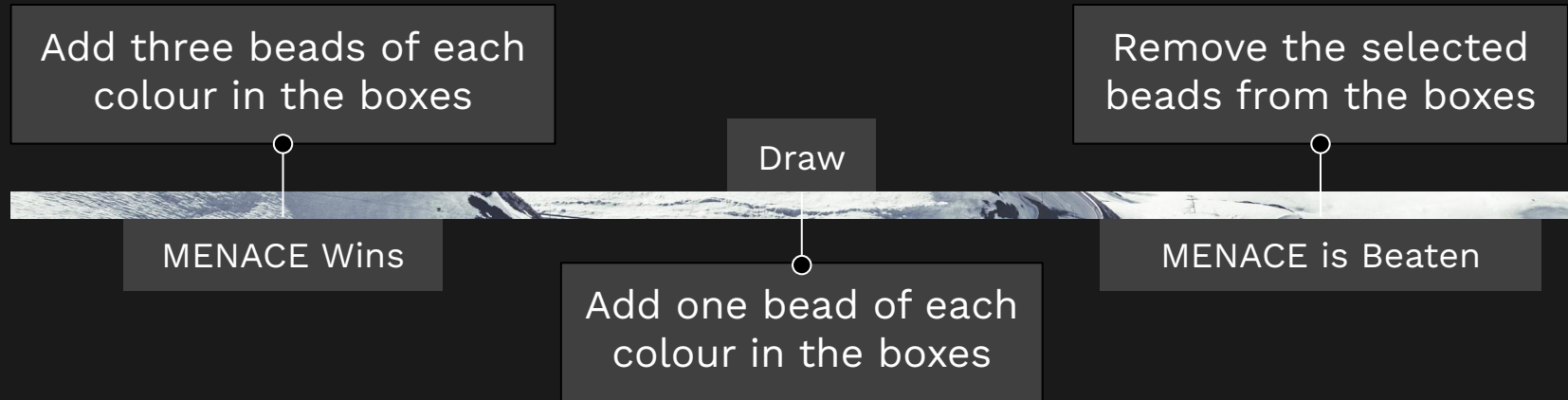



MENACE Loses



Learning

At the end of the game, it is time for MENACE to learn. Based on the outcome of the game, the following actions are performed:

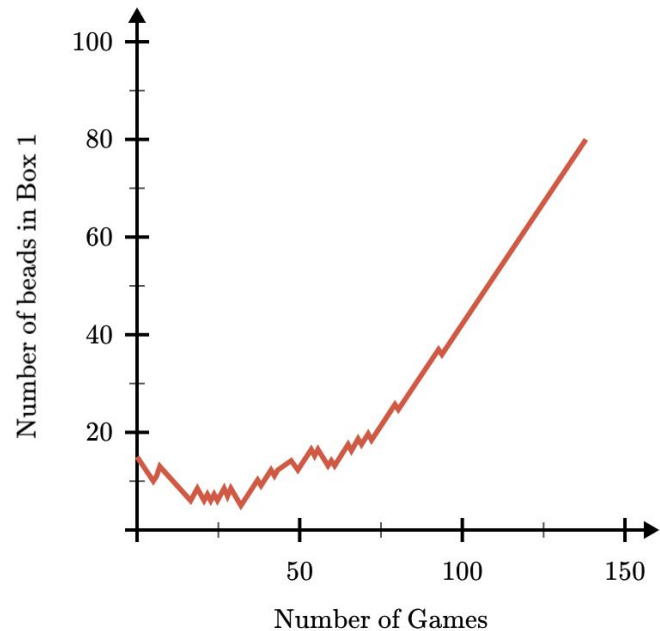




Here is a chart tracking the number of beads in the first matchbox over 140 games between MENACE and a good human player.

At first, the number of beads decreases, indicating that MENACE loses most games.

Remember that we throw away beads whenever MENACE fails.





References

- <http://people.csail.mit.edu/brooks/idos/matchbox.pdf>
- <https://www.msccroggs.co.uk/menace/>
- https://youtu.be/R9c-_neaxeU
- <https://github.com/andrewmccarthy/menace/blob/master/menace.py>



Thank you

