Project Proposal - Intro to AI

> Eliav Opatowsky - 326454550
> Liam Mohr - 209734268
> Eyal Weintroub - 327693024
> Group 66

Problem

We chose to solve the game of hanabi.
The Wikipedia page of the game explains it better, but we will summarize the rules shortly:

Summary

- Player's are aware of each other cards, but not their own (similar to blind's man bluff).
- The goal is to play a series of cards in a specific order.

Setup

- The deck contains cards in 5 suits (5 colors).
- The deck contains five numbers, 1 through 5 with:
  three 1's of each color, and two of 2,3,4,5, of each color.

Gameplay

- The players can use their turn to do one of 3 moves:
  - Give information to another player.
  - Discard a card to a discard pile.
  - Play a card from their cards.
- The information they give each other gives partial information about their cards.

Game End

- The game ends in a win when all 5's are played.

- The game ends in a loss when all fuse tokens are used up.
- Else, the game continues until the deck is used up. In that case, the highest value of each suit is summed up to the score of the game.

Solution

Our solution will be devided to 2 parts:

- An Expectimax solution
- A multi agent POMDP solution.

We chose Expectimax, because The agent does not always have the ability to perceive the entire information of the game, and thus have to make randomized decisions. Expectimax matches this modeling of the game.

We chose POMDP because the board can be modeled to a partial Markov decision process. We don't know the exact state of the board every time, and each move we make might do different actions, based on how we perceive the board.

Previous Work

We do not know of anyone who took this course, that have tried to solve this game.

There have been several research papers and solutions posted online about the game.

For example, meta wrote a paper about the game. The paper proposed a search method, which was used to improve the existing RL solution.

From what we saw, the state-of-the-art solution to this problem was Meta's one. They used tools that we haven't learned in this course (deep RL for example). That differs their solution from the one we will propose.

Meta achieved an average score of 24.68, which is a really good score, that we think will be an upper bound for our average score.

We will implement simpler solutions than they did, and try a different approach then they did (POMDP).

Evaluation

The game has a score metric, and we can easily use this score to evaluate our average performance.

Extra

We are three Talpiot cadets, and 2 of us leave shortly (26.7.24) to an officer’s course that will take much of our attention, and will end at around 11.11.24.

We are prepared to work hard on the project, and to finish it in the next 10 days, but kindly ask for consideration due to our problematic status.