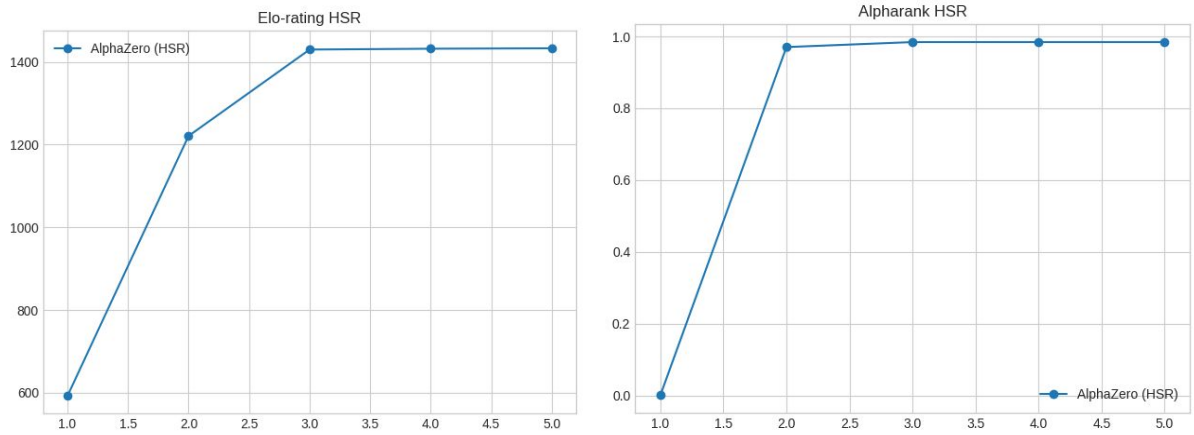
Evaluation of HSR gameplay using AlphaZero:

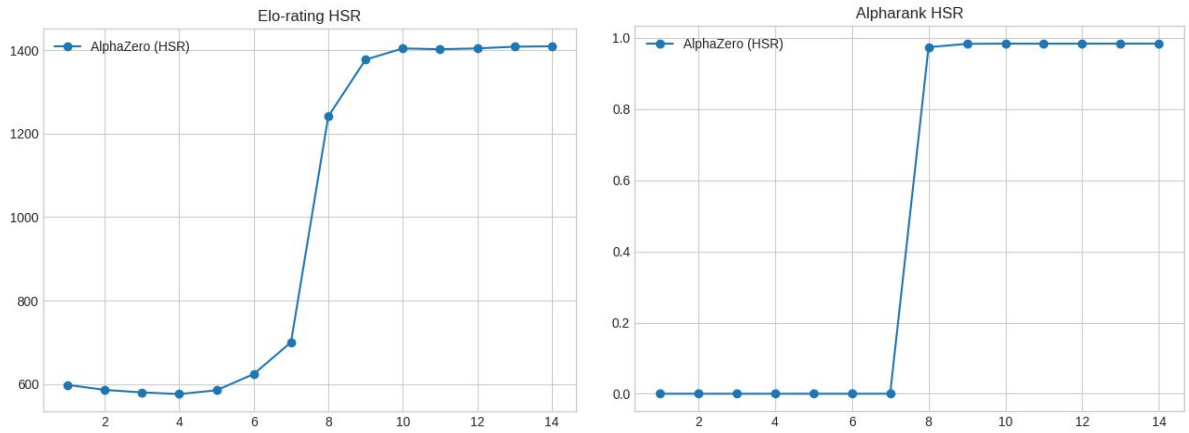
In this particular experiment, we will be using Deepmind's implementation of AlphaZero from the OpenSpiel framework. The experiment will be conducted on the HSR game and the evaluation metric used will be the Elo-rating, which is one of the most widely used techniques for calculating the relative skill of the players. We will also be using the Alpharank algorithm (α-Rank: Multi-Agent Evaluation by Evolution, Omidshafiei et. al.) which is an evaluation metric that supports both single-population (symmetric) and multi-population games. We will be using Openspiel's implementation of Alpharank in which games can be specified via Payoff-tables (or Tensors for the >2 players case) as well as Heuristic payoff-tables (HPTs). Alpharank is an evolutionary dynamics methodology for evaluating and rating agents in large-scale multi-agent interactions using Markov-Conley chains (MCCs).
For the first experiment, we use the HSR board of configuration (3,3,8) which means that we will have 3 jars, 3 tests and, 8 rungs. Based on this board we know from Bernoulli's triangle that a solution exists for this board such that the Proponent will always win the game. We run our experiments with AlphaZero having a neural network with 2 hidden layers having 128 nodes each. We use 20 as the maximum simulations parameter of the MCTS for the Alpharank algorithm. Figure() shows the results that we find for Elo-rating and the Alpharank algorithm for HSR(3,3,8).




Here we are only plotting the Elo-rating and the Alpharank of the Proponent as we would like to show how well AlphaZero performs and how quickly it can achieve the optimum strategy. As we had expected, the convergence happened very quickly as the board size is small. We can see that just after the first step of training, the Proponent starts performing very well and then maintains a constant Elo score of approximately 1450 for the rest of the training. On further looking at the training data, we observed that the Proponents consistently follows a binary search trajectory for winning games against the Opponent. The Alpharank score also shows a very similar pattern, where the score is around 0.02 after the first step of training and shoots to 0.97 just after the second step and then maintain and constant 0.99 thereafter.

In the second experiment, we are using a bigger board HSR (4,4,16) to test our algorithm under a bigger state space. This configuration of the HSR board also has a solution where the

Proponent can consistently win the game according to the Bernoulli's triangle. Since the state space of the game has increased for this experiment we use a bigger network for this evaluation with 4 hidden layers having 128 nodes in each layer. The maximum simulations parameter is set to 40 simulations for this experiment. Figure() shows the results that we find for Elo-rating and Alpharank algorithm for HSR(4,4,16).




As we can see, the Elo-rating starts at 600 and drops until step 4, after the fourth step the increase in the Elo-rating is almost linear up to step 7 after which there is a huge jump of around 550 points to step 8 and a 220 point jump to step 9  stays constant at around 1420 thereafter. Alpharank for this experiment stays constant at 0.02 for the first 7 steps and again shoots to 0.97 at step 8 and remains constant at 0.99 thereafter. The trajectory data for this experiment too shows a binary search trajectory being followed by the Proponent as the optimal strategy. Thus in both the experiments, we can see a clear phase transition taking place where the algorithm is performing poorly to one step and it achieves the optimal strategy at a certain step and maintains the same strategy thereafter.