Suarez, Trotter, Gunter

CSCI 523 Artificial Intelligence

Dr. Wilkins

Fall 2015

AI Conniption Results Report

In order to generate the data to be analyzed, the (main.py) program received the input of two heuristics to be analyzed, keeping the order as player one and player two, and wrote the log file of the game into a .pkl file (pickle is a python library that allows you to write data in binary form and decoded later to improve efficiency and space). In our first test run we used 4 different evaluation functions [Hybrid, Cells, Sols, Random] and alternated each other between player one and player two. Each game in that 4*4 grid was played 100 times, for a total of 1,600 games. This data is going to be analyzed as Test Run 1.

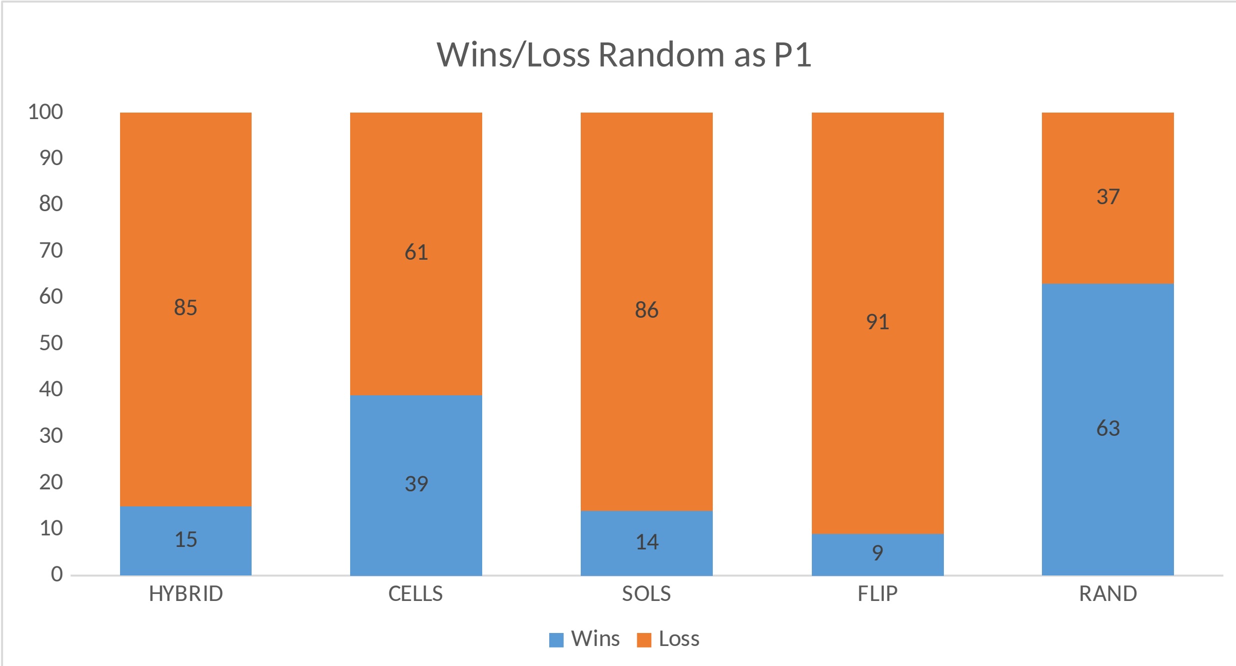
After analyzing that data we realized an improvement could be made in order to improve the behavior of the AI when playing as Player 2. A new evaluation function, Flip, was created and a new test run with now a 5 * 5 grid and a total of 2,500 games were played. The implementation and the details will be further discussed in Test Run 2.

All the presented information can be further explored in the (resilts_conniption.xlsx) file. Each single table or graph is present in its own excel workbook. Per category the lowest values are highlighted with a green background and the higher values are with a red one. The more intense the color is, the better the value is in it respective range.

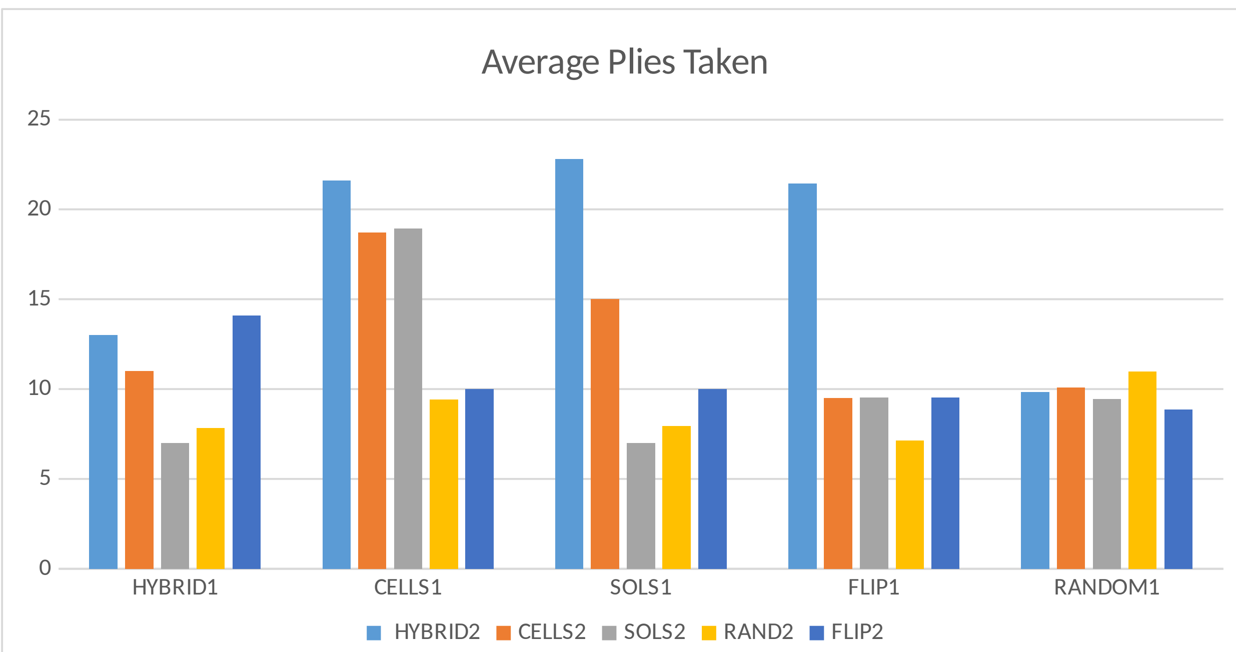
1) Test Run 1

During testing and because of the nature of the game, we realized that starting as player one provides a big advantage in terms of the final outcome of winning. The evaluation functions won all the hundred game simulations against the other evaluation functions, excluding random, in all hundred games it played against the other. Random was able to beat the other three evaluation functions playing as P2 10 times out of 300 games representing 3.33 % of the times. This behavior can be explained by the Minimax algorithm assumption that the opponent moves will always be optimal and that the opponent behavior should be similar to its own, however our random player never quite follow this rule making moves that could be considered as ‘stupid’ or meaningless.

This specific behavior previously described can be seen where Random played as P1 against the other heuristics. As can be observed in the following graph, with the observed advantage of placing first Random was able to win 77 games out of 400 for and average of 19% of the total times.



Since our three main evaluation functions [Hybrid, Sols, Cells] won a 100% of the times playing against each other as P1, we decided to test which evaluation function behaves better in overall by taking into considerations the number of chips placed on the board. The assumption is simple, the higher the number of chips placed the better the evaluation function is as P2, making it harder for P1 to beat him. The results can be observed in the following graph:

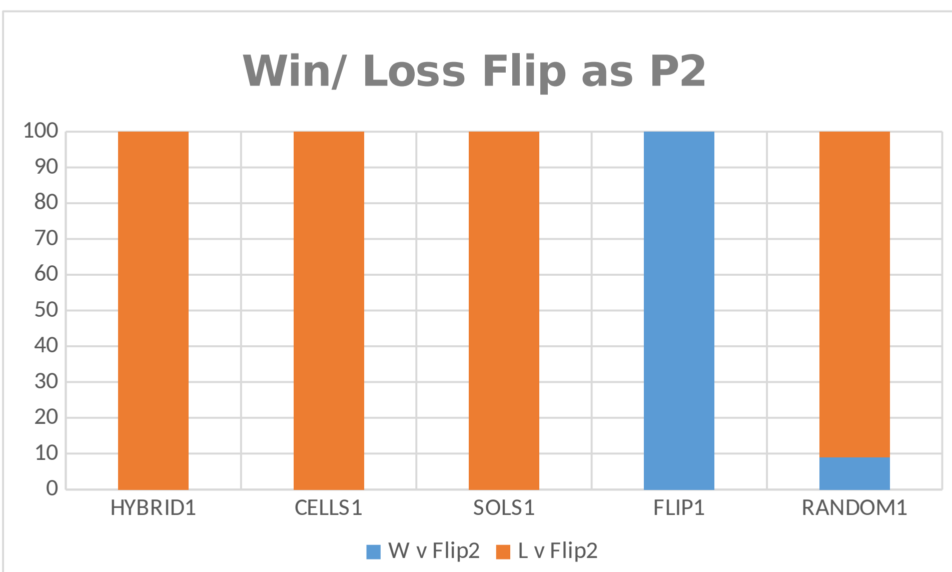


The data presented above is the average of the number of plies taken across a 100 games. The game took an average of 12.03 moves across the 1,600 games. The Hybrid playing as P2 recorded a total of 17.73 moves per play, being the best under the criteria previously analyzed. The lowest as expected was random with 8.67 and all the other ranged in between 10 and 12.

This data confirmed our assumption of having Hybrid as our best overall evaluation function. However, when playing human versus AI, and the AI playing as P2 it was fairly easy for the human to win due to a particular behavior of aggressive flipping. Our program did not place any specific weight to performing a flip, instead it analyzed all the possible moves and if the board with a flip presented a better option than the current one it will pick it every time. This was significant due to the fact that all the flips where performed early in the game and even in situation in which it might not have been entirely necessary. This lead us onto experimenting with a new evaluation function, which would include a weight cost to flipping and with the specific goal of being able to win against the previous evaluation functions a minimum but significant amount of the times.

2) Test Run 2

In this test run we created a new heuristic called Flip, which takes the scored obtained by Hybrid and adds a bias to that score by calculating the number of flips left. This new behavior creates makes the AI to perform a flip when it provides an advantage higher than the placed bias. After running the new set of tests Flip behaved as expected as P1 and won all of its 500 games. Now let's analyze how it behaved playing as P2, and how well it accomplished the main purpose of its creation.



The above table show the Win/Loss ratio of flip performing as P2. The results as can be observed are surprising. The Flip heuristic was able to the other heuristics, excluding itself, 391 out of 400 times for a 97%. Those nine losses came against random. Digging more into the data is even more surprising to be that all those victories as P2 came also with the second lowest average of plies played with 10.5.

With this data we can reach two final conclusions:

- Under similar game strategies (heuristics) the player one possess a great advantage and is ensured to win with optimal moves.
- The flipping behavior is a fundamental part of the strategy and dominating this as player one will almost guarantee a win. Flip as P1 has not be defeated by other heuristics or humans up to date. Undefeated 550-0.

3) Improving the Game Fairness

After realizing the heavy advantage the P1 possess we experimented by alternating the number of flip allowed to P1 and P2. We concluded that the closeted to a fair scenario is to allocate zero flips to P1 and one flip to P2 at the start of the game. Under this specific scenario, P2 was able to beat P1 when pitting the Flip heuristic against itself between 30% and 50% of the time.