CZ4046 INTELLIGENT AGENTS

1. Preliminary

First, we take a look at the pay-off grid, which is shown below with 0 representing cooperation and 1 as defect:

Player	Opponent1	Opponent2	Payoff
0	0	0	6
0	0	1	3
0	1	0	3
0	1	1	0
1	0	0	8
1	0	1	5
1	1	0	5
1	1	1	2

If we consider 1 round three player game, the dominant strategy will be to defect, because defecting in all scenarios offers a higher reward than cooperation. For example, given that two opponents cooperate, if the player defects, they will get 8 as opposed to 6 when they cooperated; if either one of the opponents defect, then player defect gives 5 compared to cooperation's 3; if both opponent defect, player defect will give 2 instead of cooperation's 0. If the player assume that opponents are equally rational- meaning that they will choose the dominant strategy- then everyone will defect, giving everyone (2,2,2) reward.

If we play for longer time, it then become beneficial to cooperate. This is because the player can analyse the other agents and agree that (6,6,6) offers the best reward. However, if the number of rounds is finite and known, then the player will defect at the last round knowing there will be no retaliation. By induction they will always defect.

Since the game code says it is around 90-110 rounds, we can expect to defect if we know the current number of the round is 110. The best strategy seems to be an extension of "T4T"- the player will cooperate only if the other two cooperate. Because if only one cooperate and the other one defects, the player defect will trump the player cooperate scenario.

2. Improved T4T

It is with this idea I designed the Improved T4T, which will cooperate only if both the opponents cooperate in the next round. It will give a good result as shown below, but other times T4T trumps it.

```
ImprovedTicForTacPlayer: 176.81462 points.
TolerantPlayer: 171.55148 points.
T4TPlayer: 163.2507 points.
NicePlayer: 160.30824 points.
FreakyPlayer: 160.07697 points.
RandomPlayer: 137.66695 points.
NastyPlayer: 136.91353 points.
```

Figure 1-ImprovedT4T

3. TolerantPlayer

It also seems that it is natural to choose the TolerantPlayer- but I want to test how much tolerance is good. I made TolerantPlayer40 and TolerantPlayer60 respectively- the former means 40% of cooperation is required for the player to cooperate and the latter requires 60%. Neither of them performs as well as the TolerantPlayer itself, so it seems TolerantPlayer's threshold is optimal.

```
TolerantPlayer: 195.8086 points.
T4TPlayer: 194.71478 points.
NicePlayer: 190.75143 points.
FreakyPlayer: 189.01167 points.
NastyPlayer: 170.18178 points.
RandomPlayer: 169.40488 points.
TolerantPlayer40: 161.45795 points.
TolerantPlayer60: 156.21115 points.
```

Figure 2-TolerantPlayer40 & TolerantPlayer60

4. CautiousPlayer

I also made a CautiousPlayer, and they are only tolerant when both opponents have more cooperations than defect. It is based on the TolerantPlayer but the criteria for cooperation is more stringent. The CautiousPlayer is about on par with the tolerant player- sometimes even better, but it is unclear to see which is the best.

```
CautiousPlayer: 170.70636 points.
T4TPlayer: 164.73318 points.
TolerantPlayer: 157.9055 points.
NicePlayer: 154.45526 points.
NastyPlayer: 143.57423 points.
FreakyPlayer: 138.71558 points.
RandomPlayer: 133.62506 points.
```

Figure 3-CautiousPlayer

5. UtilityPlayer

Then I made the UtilityPlayer, which calculate the expected reward given histories of the opponents. The formula is as follows:

$$U(a) = P_{1c}P_{2c}Payoff(c,c,a) + P_{1d}P_{2c}Payoff(d,c,a) + P_{1c}P_{2d}Payoff(c,d,a) + P_{1d}P_{2d}Payoff(d,d,a)$$

Where a represents action $\in$ {cooperate,defect} and P_{1d} means the probability that 1 defects. But this is not the best as it will tend to defect when both opponents are playing nice to get the high reward and causing a retaliation in the next round. Then it will keep defecting since both opponents defect a lot, and player defect gives the best utility.

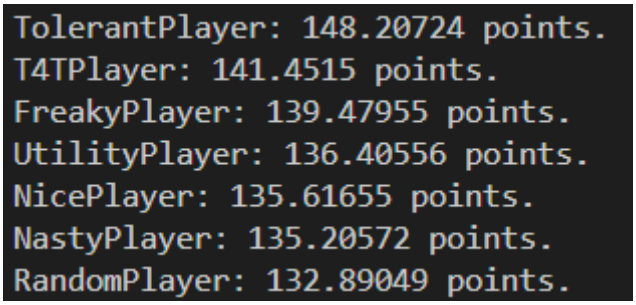

Figure 4-UtilityPlayer

6. StochasticPlayer

But if we modify the utility function to add the future utility inside. U(a) will be the sum of the current utility of this round and all the future utility. However, to do this, we need to know the "vengeance" and "niceness" of the opponent- namely, if we defect, will opponent defect immediately? If so, the player should probably not defect. We do that by summing the number of defects immediately after our defect, then divided by the number of my defect. We perform the same to niceness scale and multiply the two numbers to obtain a one-step look ahead. Assuming we defect this round, the look ahead will be:

$$U'(d) = (1-V_1)(1-V_2)P_{1c}P_{1c}P_{2c}Payoff(c,c,d) + V_1(1-V_2)P_{1d}P_{2c}Payoff(d,c,d) \\ + (1-V_1)V_2P_{1c}P_{2d}Payoff(c,d,d) + V_1V_2P_{1d}P_{2d}Payoff(d,d,d)$$

In this equation, V_1 means the vengeance score of Opponent 1. Similarly, we can do the same for one step look ahead for U'(c). By summing them up with the original payoff, we can make a StochasticPlayer that estimates the opponent's move and react accordingly. I also realise that if the last two opponents have reached an agreement, they will hardly deviate from it if you do the same. So as long as (c,c,c) and (d,d,d) are reached, the rational opponents will not move from the state. The trick is then to play along. This also gels well from the previous observation that T4T and Tolerant are the best, so incorporated such a law in my StochasticPlayer.

Although StochasticPlayer can achieve very good result as shown below, it can at times perform badly due to rough estimations.

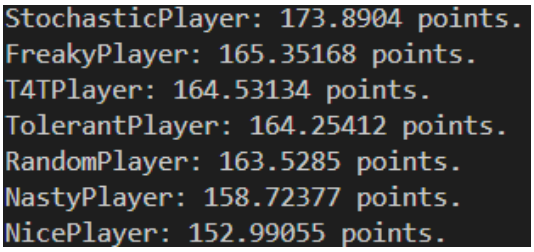

Figure 5-StochasticPlayer Performing Well

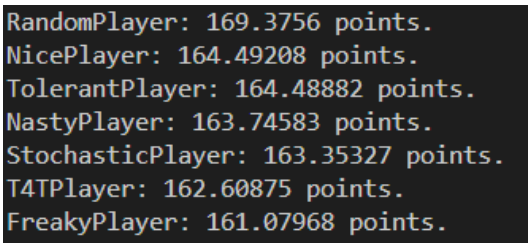

Figure 6-Stochastic Player Performing Badly

Interestingly, it also seems to enhance RnadomPlayer, since when it performs badly the RandomPlayer will perform well.

I then put all players I made plus the Joss and Tester in lecture notes and see the result:

```
Tournament Results
StochasticPlayer: 516.78265 points.
TolerantPlayer: 507.11337 points.
CautiousPlayer: 504.01373 points.
NicePlayer: 500.25336 points.
T4TPlayer: 484.38757 points.
ImprovedTicForTacPlayer: 457.71838 points.
Joss: 442.5335 points.
FreakyPlayer: 433.30695 points.
Tester: 427.74698 points.
TolerantPlayer60: 412.55338 points.
RandomPlayer: 408.53452 points.
TolerantPlayer40: 407.54865 points.
UtilityPlayer: 393.94763 points.
NastyPlayer: 369.0285 points.
```

Figure 7-Ultimate Result

7. Conclusion

In conclusion, both CautiousPlayer, ImprovedT4TPlayer and StochasticPlayer show good promise, and the original T4T and TolerantPlayer are not bad. It is inconclusive which one is the best, and more testing may need to be done. In my code, I decided to submit the StochasticPlayer for its ability to detect other agent's traits, albeit the method of detection is crude.